ENVIRONMENTAL RESEARCH
LETTERS

PURPOSE-LED PUBLISHING™

LETTER • OPEN ACCESS

Deep learning improves sub-seasonal marine heatwave forecast

To cite this article: Di Sun *et al* 2024 *Environ. Res. Lett.* **19** 064035

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
29 January 2024

REVISED
27 March 2024

ACCEPTED FOR PUBLICATION
1 May 2024

PUBLISHED
24 May 2024

LETTER

Deep learning improves sub-seasonal marine heatwave forecast

Di Sun[1,2,*], **Zhao Jing**[1,2] **and Hailong Liu**[2]

[1] Frontiers Science Center for Deep Ocean Multispheres and Earth System and Key Laboratory of Physical Oceanography, Ocean University of China, Qingdao, People's Republic of China
[2] Laoshan Laboratory, Qingdao, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: oucsundi@gmail.com

Keywords: marine heatwaves, sub-seasonal forecast, deep learning, bias correction

Supplementary material for this article is available online

Abstract

Marine heatwaves (MHWs) are extreme anomalously warm water events, which are projected to cause increasing numbers of disastrous impacts on ecosystems and economies under global ocean warming. Our ability to forecast MHWs determines what effective measures can be taken to help reduce the vulnerability of marine ecosystems and human communities. In this study, we combine a deep learning model, the convolutional neural network, with a real-time sub-seasonal to seasonal physical forecast model, improving MHW forecast skills by nearly 10% of the global average in leading two weeks by correcting the physical model bias with observational data. This improvement has a nearly consistent influence (∼10%–20%) on a global scale, reflecting the wide-coverage promotion by deep learning. This work reveals the advantages and prospects of the combination of deep learning and physical models in ocean forecasts in the future.

1. Introduction

Over the past two decades, extreme anomalously warm water events, known as marine heatwaves (MHWs), have gained increasing popularity in both scientific and public communities due to their significant destruction on marine ecosystems and economic society (Smale *et al* 2019, Viglione 2021). The impacts on marine ecosystems include mass coral bleaching (Hughes *et al* 2017, 2018), extensive mass mortality of marine organisms (Garrabou *et al* 2009, Thomson *et al* 2015), changes in distributions of fish communities (Ñiquen and Bouchon 2004, Mills *et al* 2013) and ultimately sudden and dramatic shifts in ecological structure and functioning (Wernberg *et al* 2013). The consequences of economic effects have also been documented due to the vulnerable aquaculture and economically valuable fisheries (Cavole *et al* 2016). All these disasters are projected to become more severe (Oliver *et al* 2019) due to increasing frequency, duration and intensity of MHWs under global ocean warming (Frölicher *et al* 2018, Oliver *et al* 2018, 2019, Laufkötter *et al* 2020).

To efficiently reduce the vulnerability of marine ecosystems even human communities, and take adaptive action before disasters occur, our ability to forecast MHWs plays a vital role. Several efforts have been made to investigate MHW predictability and forecast skills in both seasonal and sub-seasonal time scale (Benthuysen *et al* 2021, De Burgh-Day *et al* 2022, Jacox *et al* 2022). The predictability of global seasonal MHW forecasts has been assessed by a large multi-model ensemble of global climate forecasts with lead times of up to a year (Jacox *et al* 2022), showing regional skillful forecasts related to large-scale climate modes such as El Niño. However, because of the scarcity of extreme heatwaves, their occurrence in historical data is unprecedented. Thus, estimating their occurrence probability, building reliable statistics and resolving the high-dimensional nonlinear dynamics globally with physical models are still challenging (Jacques-Dumas *et al* 2022), especially at the shorter sub-seasonal time scale without seasonal variation or the state of large-scale climate patterns.

Recently, there have been several studies that have gone beyond the physical models, instead attempting to introduce machine learning methods to forecast MHWs, which takes the advantages of forecasting extreme events using limited information about the events' drivers. For the straightforward MHW forecast, one machine learning method, called random forest, has been adopted in the weekly forecast in the northeast Pacific (Giamalaki *et al* 2022). Even if the small- and large-scale MHWs could be efficiently

captured, the results tend to be more smoothed, i.e. lower intensity and more events with such a simple model. Forecast skills have the potential to be further improved with advanced methods. Because of this, more accurate and well-tuned deep learning models have been employed in the end-to-one tasks for both atmospheric and MHWs (Chattopadhyay *et al* 2020, Jacques-Dumas *et al* 2022). These models choose the convolutional neural network (CNN) as their architecture, receiving a whole spatial image and producing one single value, usually a region-averaged forecast result. The task of learning a single result from large amounts of data is relatively simple and thus shows higher accuracy (Ham *et al* 2019). However, a single-value forecast for a large region is of less help to some stakeholders. Aside from this, some studies open up a new pathway for forecasting of sea surface temperature (SST) followed by the subsequent detection of MHWs. The progress includes long short-term memory models for time-series prediction in a single location (Karevan and Suykens 2020, Wolff *et al* 2020) and, lately, an end-to-end seasonal SST forecast model on a global scale (Taylor and Feng 2022). Yet, as is well known, the direct forecasting of extreme events, including MHWs, using deep learning is absolutely not an easy task, attributed to the need to learn from sparse events and class imbalance. The strong persistence in the ocean extension periods also makes it difficult to forecast MHWs in a subseasonal time scale. So far, there have been few end-to-end deep learning models for global MHW subseasonal forecasts.

In this study, instead of directly using a single deep learning model to forecast MHWs, we employ a CNN model combined with a real-time sub-seasonal to seasonal (S2S) physical forecast model, improving MHW forecast skills globally in the leading 14 days by correcting the bias of the physical model with observational SST as a calibration factor. The results from directly forecasting MHWs using a single deep learning model are also compared. We choose a sub-seasonal scale because the accumulated thermal stress of an MHW event can play a role for several days in marine ecosystems, and ocean stakeholders could benefit from such short-term forecast. After CNN training, the forecast skills have been significantly elevated compared with the original physical model on test data, which reveals the advantages and prospects of the combination of deep learning and physical models in ocean forecasts in the future.

2. Materials and methods

2.1. Data

Real-time forecast products and observational products are used to train our CNN model as predictors and the ground truth, respectively. Due to the availability of spatio-temporal resolution, S2S SST forecast products from UKMO-GloSea (United Kingdom Meteorological Office coupled Global Seasonal forecast, abbreviated as UKMO hereinafter) are archived in the leading 14 days on a 1.5° × 1.5° field from 2016–2022 with daily resolution (Vitart *et al* 2017). In the previous studies related to SST, UKMO has been evaluated reasonably well against the European Centre for Medium-Range Weather Forecasts (ECMWF) (Vitart 2017, Qin *et al* 2020, 2022), a widely recognized product but with lower temporal resolution, excluded by this study. The observational products are obtained from the National Oceanic and Atmospheric Administration Optimum Interpolation SST (NOAA OISST) version 2.0 daily data (Reynolds *et al* 2007) and smoothed onto an identical 1.5° × 1.5° field from 1982–2022, which have been widely used in previous MHW analysis (Hobday *et al* 2016, Oliver *et al* 2017, 2021, Holbrook *et al* 2019, Sen Gupta *et al* 2020).

We identify MHWs globally from 0–360° E, 60° S–60° N and produce binary images from OISST to generate the ground truth. Following Hobday *et al* (2016), an MHW event is defined as a discrete anomalously warm water event. Here, the term 'discrete' means that an MHW event should have well-defined start and end times. 'Anomalously warm' is defined by reference to a baseline climatology, i.e. a seasonally varying threshold from 1982–2011, given by OISST here. SST anomalies (SSTA) above the baseline climatology of the leading 14 days from UKMO and the past day from OISST are used as predictors. We remark that we use the SSTA of the past day from observations because today we only know yesterday's information. Explicitly, we denote the current day as t_0, the SSTA of t_{-1} from OISST is considered as the calibration factor and $t_1, \ldots, t_{14}$ from UKMO are concatenated as input data for our CNN model. Corresponding true MHW events (labeled as 1, and 0 otherwise) of $t_1, \ldots, t_{14}$, detected by OISST, are used for the ground truth when training our CNN model. The land grids in both input and output data are masked as zero to avoid infinite values during the gradient descent. This setting does not affect the detection of MHWs because SSTAs are usually greater than zero where MHWs occur.

2.2. CNN architecture

We establish the state-of-the-art CNN model (named as 'MHWUNet' hereinafter) based on a U-Net backbone, which has been well adopted in image segmentation tasks in diverse fields, such as imaging diagnosis in biomedical science, eddy detection and sea ice forecasting from satellite remote sensing images (Ronneberger *et al* 2015, Lguensat *et al* 2018, Andersson *et al* 2021). The term 'U-Net' comes from its U-shaped network structure, which includes an encoder (down-sampling path on the left) and a decoder (up-sampling path on the right) module. Due to the skip connection and feature fusion between down-sampling and up-sampling, U-Net is

capable of capturing and retaining information about the input data, and thus has been maintaining its growing popularity in different kinds of end-to-end application tasks. Given the potential overfitting in the simple feature of SSTAs, the powerful residual network block (He *et al* 2016), designed to solve gradient vanishing and degradation problems with increased layers, is introduced into each down-sampling and up-sampling block of MHWUNet. The use of a residual block ensures that a very deep network can be trained. Furthermore, to enhance the performance of the MHWUNet model, an efficient attention module, named the convolutional block attention module (CBAM) is integrated into the central block of the U-Net backbone. CBAM includes one channel attention module and one spatial attention module to capture and highlight the important features behind different channels and locations within feature maps (Woo *et al* 2018). It may play a role by focusing on SSTAs of different days (by channel attention) and MHWs with different spatial extent scattered around the global ocean (by spatial attention) in MHWUNet.

Figure 1 shows the MHWUNet architecture used in this study, which contains one central block (CBAM), three-times down-sampling and identical three-times up-sampling blocks. In each encoder block, we use two convolutional layers with a 3×3 kernel size and a fixed channel size of 32, which is selected from various attempts behind batch normalization layers, followed by the rectified linear unit (ReLU) layer as a nonlinear activation function and a 2×2 maximum pooling layer. A dropout layer is included after the activation layer to avoid overfitting. A residual block in each block is inserted between the input feature and output feature after two-time convolution. In each decoder block, a skip connection combines the feature maps of the encoder with the up-sampling features of the decoder, after which the same convolutional layers as the encoder are adopted. For the point-wise classification task, a convolutional layer combined with a sigmoid activation function is added in the output layer. Numerous attempts and hyperparameter tuning lead to this architecture instead of the preliminary model of UNet (Ronneberger *et al* 2015).

The objective of the MHWUNet model is to minimize the loss function defined as weighted dice loss, which has been widely used in segmentation tasks (Milletari *et al* 2016, Lguensat *et al* 2018) following

$$\text{loss} = 1 - \left(\alpha_1 \times \frac{2\left|P_{\text{mhw}} \cap G_{\text{mhw}}\right|}{\left|P_{\text{mhw}}\right| + \left|G_{\text{mhw}}\right|} + \alpha_2 \times \frac{2\left|P_{\text{bg}} \cap G_{\text{bg}}\right|}{\left|P_{\text{bg}}\right| + \left|G_{\text{bg}}\right|} \right), \tag{1}$$

where P_{mhw} represents the predicted region of MHWs, G_{mhw} represents the ground truth region of MHWs, P_{bg} represents the predicted region of the background (non-MHWs), G_{bg} represents the ground truth region of the background, $|P|$ and $|G|$ are denoted as the sum of elements in each area, and $\cap$ represents the sum of overlapped elements of two images. The coefficients of α_1 and α_2 are tuning parameters here for class imbalance between MHWs and the background following

$$\alpha_1 = \frac{n_1}{n_1 + n_2}, \alpha_2 = \frac{n_2}{n_1 + n_2}, \tag{2}$$

where n_1 and n_2 represent the pixel numbers of non-MHW and MHW, and α_1 and α_2 represent the coefficients of MHW and non-MHW. In this way, the rare class will have a larger coefficient to maintain class balance and $\alpha_1 = 0.83$, $\alpha_2 = 0.17$ is used in this study (sensitive experiments are shown in the supporting information figure S1). We caution readers that these two hyper parameters may vary from different thresholds for MHW definition and sensitive experiments need to be conducted. To update the parameters in the CNN model, we use an Adam optimizer (Kingma and Ba 2014) with an initial learning rate of 10^{-3}, reducing in half with patience of 10 epochs. We set 400 training epochs in total to train the MHWUNet model with an early stopping strategy for 80 epochs. The model is trained with a batch size of 4 on a single Tesla V100 GPU on a random 80% for the input data from 2016–2022 and tested on the last 20% of the data. Another division using the first 80% as training data and testing on the last 20% is also tested and added in the supporting information (figure S2) to exclude the potential influence of unseen data. The supporting information (figure S3) shows the evolution of the loss function during the MHWUNet training process.

2.3. Model evaluation

The evaluation metrics used in this study are the statistical scores from the confusion matrix, including the F_1 score (F_1) and Matthews correlation coefficient (*MCC*), which have been widely leveraged to evaluate the skills of the binary forecast.

The confusion matrix presents the proportions of the actual versus predicted values for whether there is an MHW or not, from which the P_r (named as the precision) of statistical scores is defined as follows:

$$P_r = \frac{TP}{TP + FP}, \tag{3}$$

where *TP* (referenced to true positive) represents the sample predicted as 1 with label 1, and *FP* (referenced to false positive) represents the sample predicted as 1 with label 0. Therefore, P_r refers to how many of the samples we predict as True are actually True. The other score R_c (named as recall) reflects the accuracy rate, i.e. how many samples that are actually True are picked out following

$$R_c = \frac{TP}{TP + FN}, \tag{4}$$

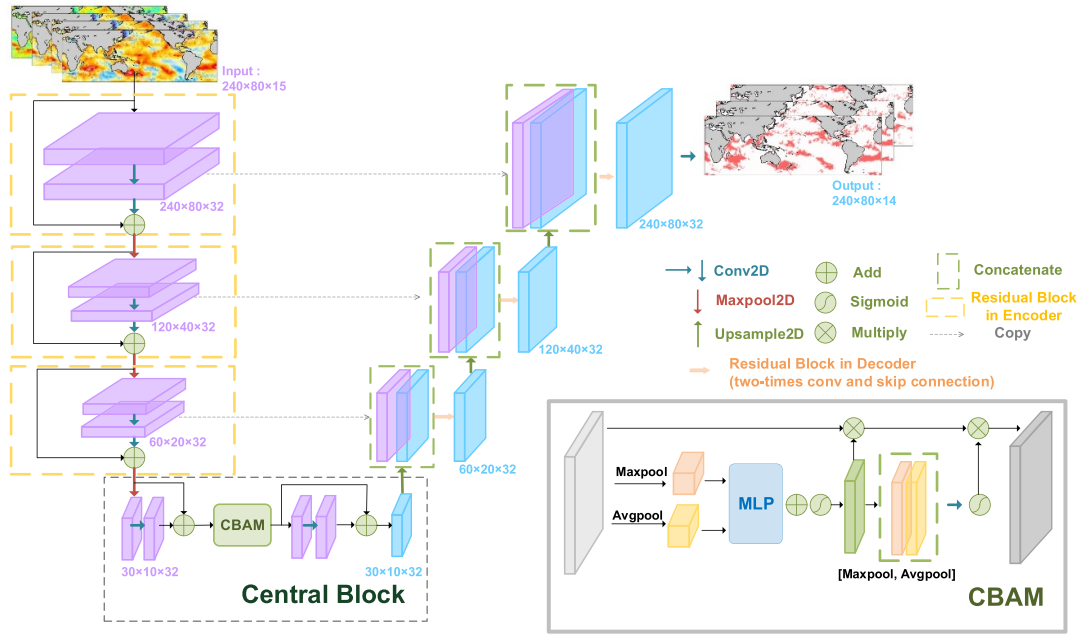

Figure 1. A summary of the architecture of the MHWUNet. It receives leading 14-day $(t_1, \ldots, t_{14})$ SSTAs from UKMO and a calibration SSTA from OISST at the last channel as input. The outputs are 14-day forecast $(t_1, \ldots, t_{14})$ of binary images (MHW: 1, non-MHW:0). The number in the bottom right of the convolutional layers denotes the resolution and number of feature maps.

where *FN* (referenced to false negative) represents the sample predicted as 0 with label 0. Combining both P_r and R_c leads to the comprehensive score F_1 following

$$F_1 = \frac{2 \times P_r \times R_c}{P_r + R_c}. \tag{5}$$

Taking into account the four classification results in the confusion matrix, *MCC* is a more comprehensive score that can provide a more accurate performance evaluation in imbalanced data following

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FN)(TP+FP)(TN+FP)(TN+FN)}}. \tag{6}$$

The maximum of the *MCC* reaches 1, and a score above 0 indicates the forecast better than random chance.

3. Results

3.1. Improvement in the forecast skills

Forecast skill scores of leading 14 days on the test data are shown in figure 2, which highlights the advantages of CNN bias correction for sub-seasonal MHW forecasts on a global-average scale. With regard to the statistical skill scores in figure 2, both F_1 and the *MCC* from MHWUNet show higher skills than the uncorrected UKMO model. Specifically, the F_1 score of the MHWUNet shows 7.8% improvement on a global-average scale on the leading 1st day, and 6.8% and 8.3% improvement on the leading 7th day and 14th day, respectively, compared to the S2S model UKMO with a benchmark of OISST (orange line and blue line in figure 2(a)). The improvement in the *MCC* is 9.4%, 8.3% and 10.2% on the leading 1st day, 7th day and 14th day, respectively (orange line and blue line in figure 2(b)). This means bias correction by MHWUNet persists with its effectiveness when the forecast time increases, and the longer input length has the potential to maintain this effectiveness (supporting information figure S4). The results also show that there is a gap for the physical forecast models to overstep the persistence ($P_{t+1} = P_t$, where P represents the predicted values) in the ocean extension periods due to the model configuration with a self-consistent initial field. However, when the forecast time exceeds two days, the corrected results from our MHWUNet perform obviously better than the persistence from OISST, which decreases faster as time goes by (orange line and purple line).

Another pathway to forecast MHWs is to use the observations only. We use SSTAs from OISST of the past 14 days as input and the binary images of MHWs from OISST in the leading 14 days as output to train ForeUNet. The model architecture is the same as MHWUNet. The forecast skills of the global mean also show a large improvement compared with UKMO (green lines in figure 2), i.e. 5.9% improvement in F_1 and 7.9% improvement in the *MCC* on the leading 14th day. However, as we will illustrate in section 3.2, the spatial patterns in the forecast days of ForeUNet tend to imitate the patterns of the first day, showing almost unchanged patterns.

3.2. Global patterns of the forecast skills

Spatial patterns for MHWUNet-corrected results are shown in figure 3. In comparison with the uncorrected UKMO model, the results from MHWUNet

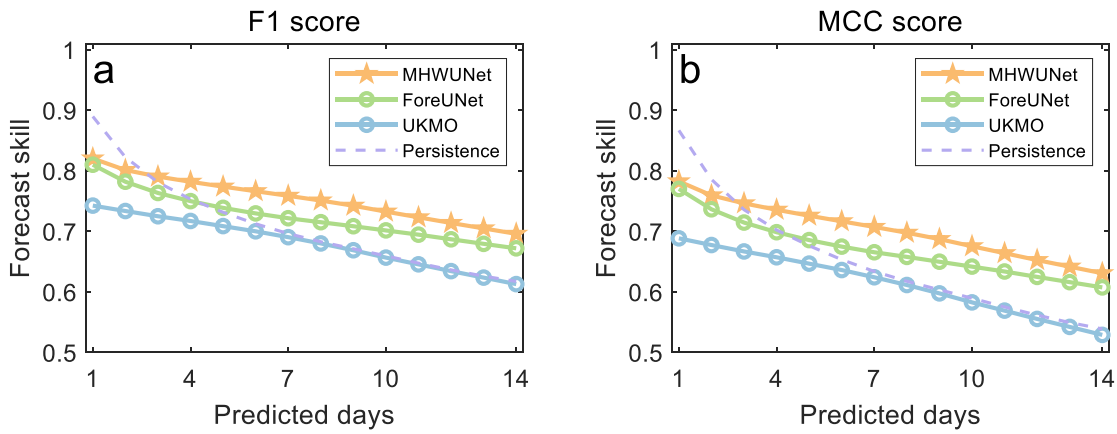

Figure 2. Skill scores of the evaluation metrics for leading 14 days from UKMO, MHWUNet, ForeUNet and the persistence from OISST. (a) The statistical skill score F_1 for MHWUNet correction (in orange), ForeUNet (in green), the uncorrected UKMO model (in blue) and the persistence with OISST (in purple). (b) Same as (a) but for the statistical skill score *MCC*.

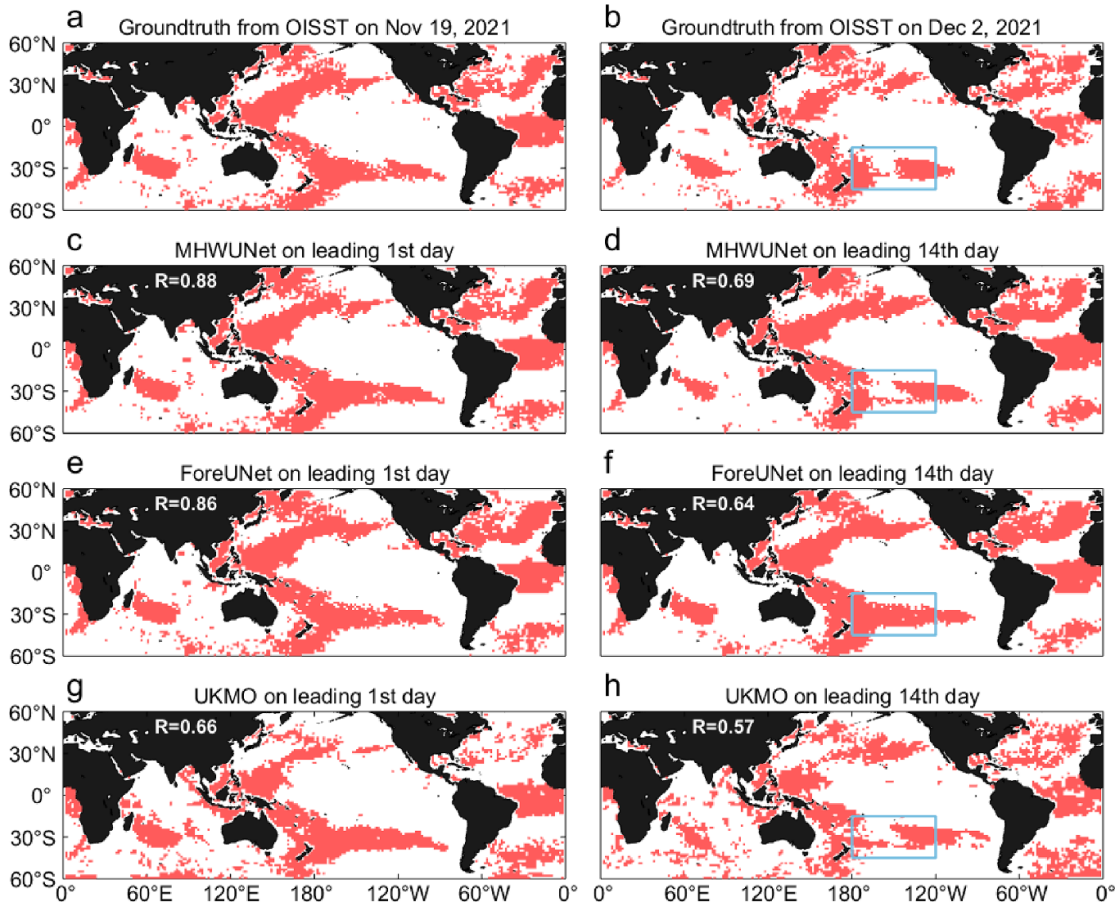

Figure 3. Spatial patterns for a case of MHWUNet correction. (a) Groundtruth from OISST on 19 November 2021. (b) Groundtruth from OISST on 2 December 2021. (c) MHWUNet correction results on the corresponding leading 1st day. (d) MHWUNet correction results on the corresponding leading 14th day. (e) ForeUNet results on the corresponding leading 1st day. (f) ForeUNet results on the corresponding leading 14th day. (g) Uncorrected UKMO results on the corresponding leading 1st day. (h) Uncorrected UKMO results on the corresponding leading 14th day. Here, the detected/predicted MHWs are shown in red and otherwise in white.

show greater resemblance to OISST. This similarity is not only shown in the first few days (figures 3(a), (c) and (g)) but also lasts and is even more significant up to the leading 14th day (figures 3(b), (d) and (h)). On the leading 1st day, MHWs from the UKMO model have a relatively closer pattern to OISST with a correlation coefficient of 0.66 between UKMO and OISST, which has been improved to 0.88 with the correction of our MHWUNet. The spatial patterns in the Southern Indian Ocean, Northern Atlantic and

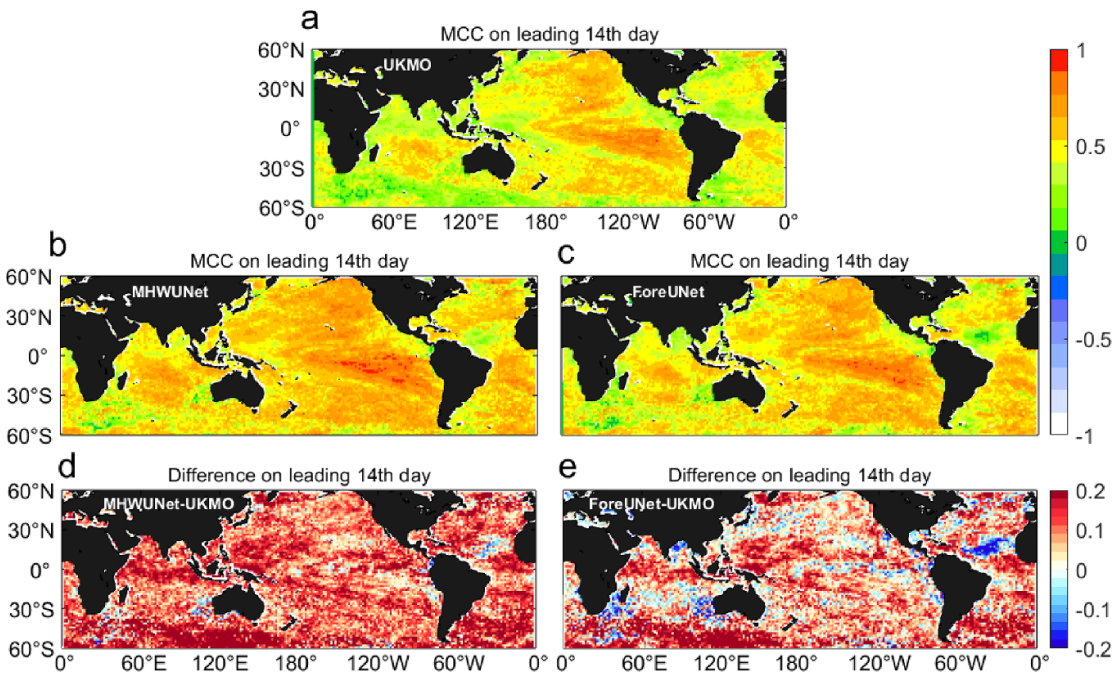

Figure 4. Spatial patterns of the statistical forecast skill improvement on the leading 14th day. (a) Spatial patterns of the MCC from the uncorrected UKMO on the leading 14th day. (b)–(c) The same as (a) but for MHWUNet and ForeUNet, respectively. (d) The difference in the MCC between MHWUNet and the UKMO model (MHWUNet—UKMO). (e) The same as (d) but for ForeUNet and UKMO (ForeUNet—UKMO).

Northeastern Pacific become more consistent after the CNN correction. On the leading 14th day, MHWs from the UKMO model exhibit large differences to OISST in the Northeastern Pacific, South Pacific and Indian Ocean with the correlation coefficient decreasing to only 0.57 between UKMO and OISST. Yet, with the correction of MHWUNet, these differences are largely erased out and the similarity to OISST reappears in an ocean-basin scale with the correlation coefficient elevated to 0.69. This increase reflects the persistent improvement in MHWUNet performance with time increasing.

Spatial patterns of ForeUNet are also shown in figure 3. However, instead of exhibiting the same pattern as OISST on the leading 14th day (figures 3(b) and (f)), ForeUNet tends to imitate the spatial patterns of the first day in the following days (figures 3(e) and (f)), showing a higher correlation coefficient between the first leading day and the 14th leading day ($R = 0.89$). But this correlation coefficient is only 0.6 for observations. This is partly because there is strong persistence in the ocean extension periods. Maintaining consistency with the previous days also results in a lower loss function, and thus there will be hardly any changes in the forecast for the next consecutive days if using a single deep learning model with only observations. Nevertheless, even if the SST is changing slowly, MHWs are ebbing and flowing frequently over a two-week time scale (Sun *et al* 2023), and the forecast results that are consistent with the first day lead to unreasonably continuous patterns. This may be partly settled down by introducing the physical model UKMO with larger variability and the observations as a calibration factor for the initial and forced fields. The blue box in figure 3 displays the detachment and separation of MHWs in the east and west parts. MHWUNet learns this state, while the results from ForeUNet maintain the same patterns as the first leading day.

To further evaluate the improvement in the forecast skills from a spatial perspective, global patterns of the statistical skill score *MCC* on the leading 14th day from MHWUNet, ForeUNet, the UKMO model and their differences are shown in figure 4 (patterns of F_1 in the supporting information figure S5). The forecast skill score *MCC* has a nearly consistent increase (∼10%–20%) on a global scale from MHWUNet (figure 4(d)), which depicts the wide-coverage improvement in the sub-seasonal MHW forecast. The highest elevation appears in the south of Australia in the South Ocean, north of the Indian Ocean and south of the Atlantic Ocean. This improvement from ForeUNet is relatively lower due to its almost unchanged spatial patterns in a two-week time scale (figure 4(**e**)).

4. Conclusions and discussion

In this study, by employing a CNN model (MHWUNet) on an S2S real-time physical forecast model (UKMO), MHW forecast skill scores have been largely and stably elevated in leading 14 days by correcting the UKMO bias with the observational SST as a calibration factor, showing ∼8% improvement in

F_1 and ∼9% improvement in the MCC in the 14-day mean on a global-average scale compared to UKMO with a benchmark of OISST. This improvement has a nearly consistent influence (∼10%–20%) on the global ocean in the leading 14 days, reflecting the wide-coverage promotion by MHWUNet correction.

MHWs are extreme anomalously warm water events with unstable deformation and displacement (Sun *et al* 2023). As is widely recognized, forecasting extreme and irregular ocean phenomena using deep learning is very challenging. Although the former could be partly settled down by introducing class balance coefficients in the loss function (i.e. the loss function used in this study), irregular shapes and the lack of time cycles in the MHW sub-seasonal forecast make it difficult for the CNN models to capture specific patterns in both space and time. More importantly, due to the strong persistence in the ocean extension periods, if only using observations from the past 14 days to predict MHWs for the next 14 days via a single deep learning model, the results tend to imitate the spatial patterns of the first leading day in the following days. In this way, the loss function can still be much lower and forecast skills can be higher. However, even if the SST is changing slowly, MHWs are ebbing and flowing frequently in a two-week time scale (Sun *et al* 2023), and the forecast results that are consistent with the first leading day lead to unreasonably continuous patterns. As illustrated in this study, a combination of the CNN model with an ocean physical model that provides the variability and observations that may correct the initial and forced fields of the physical model (supporting information figure S6) overcomes the above-mentioned difficulties and makes the high-quality sub-seasonal forecast for MHWs a reality; this highlights the collective roles from deep learning and the physical model. In addition to the direct bias correction of MHWs, the results from the other pathway, for correction of SST and subsequent detection of MHWs, have been shown in the supporting information (figure S7). This pathway makes sense as well, with a slightly lower skill with the leading time increasing.

We finally remark that the achievement uses only a single Tesla V100 GPU with inference time less than 2 s on the test data with a batch size of 16. Therefore, the use of CNN model not only improves the forecast skills but also accelerates the calculation speed, far exceeding that of the physical models with momentous numerical expense; this gives deep learning the vitality to replace the expensive parts of traditional physical models. The integration of deep learning and physical models, whether it is bias correction, module replacement or other tasks, thus provides a promising approach for application in future ocean forecasts.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This study was supported by Taishan Scholar Funds (tsqn201909052) and Fundamental Research Funds for the Central Universities (202461005). Computational resources were supported by Laoshan Laboratory. The authors declare no competing interests.

Open research

This study uses open data only. UKMO S2S real-time forecast data can be obtained from https://apps.ecmwf.int/datasets/data/s2s-realtime-daily-averaged-egrr. The observational OISST V2.0 data are available from https://climatedataguide.ucar.edu/climate-data/sst-data-noaa-optimal-interpolation-oi-sst-analysis-version-2-oisstv2-1x1. The CNN architecture developed in this study is programmed using Tensorflow 2.3.0 by Python 3.6.5 and has been uploaded to https://github.com/cindyisok/MHWUNet.

References

Andersson T R *et al* 2021 Seasonal Arctic sea ice forecasting with probabilistic deep learning *Nat. Commun.* **12** 1–12

Benthuysen J A, Smith G A, Spillman C M and Steinberg C R 2021 Subseasonal prediction of the 2020 Great Barrier Reef and Coral Sea marine heatwave *Environ. Res. Lett.* **16** 124050

Cavole L M *et al* 2016 Biological impacts of the 2013–2015 warm-water anomaly in the northeast Pacific: winners, losers, and the future *Oceanography* **29** 273–85

Chattopadhyay A, Nabizadeh E and Hassanzadeh P 2020 Analog forecasting of extreme-causing weather patterns using deep learning *J. Adv. Model. Earth Syst.* **12** e2019MS001958

De Burgh-Day C O, Spillman C M, Smith G and Stevens C L 2022 Forecasting extreme marine heat events in key aquaculture regions around New Zealand *J. South. Hemisph. Earth Syst. Sci.* **72** 58–72

Frölicher T L, Fischer E M and Gruber N 2018 Marine heatwaves under global warming *Nature* **560** 360–4

Garrabou J *et al* 2009 Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave *Glob. Change Biol.* **15** 1090–103

Giamalaki K, Beaulieu C and Prochaska J X 2022 Assessing predictability of marine heatwaves with random forests *Geophys. Res. Lett.* **49** e2022GL099069

Ham Y G, Kim J H and Luo J J 2019 Deep learning for multi-year ENSO forecasts *Nature* **573** 568–72

He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8

Hobday A J *et al* 2016 A hierarchical approach to defining marine heatwaves *Prog. Oceanogr.* **141** 227–38

Holbrook N J *et al* 2019 A global assessment of marine heatwaves and their drivers *Nat. Commun.* **10** 2624

Hughes T P et al 2017 Global warming and recurrent mass bleaching of corals Nature 543 373–7
Hughes T P et al 2018 Spatial and temporal patterns of mass bleaching of corals in the Anthropocene Science 359 80–83
Jacox M G, Alexander M A, Amaya D, Becker E, Bograd S J, Brodie S, Hazen E L, Pozo Buil M and Tommasi D 2022 Global seasonal forecasts of marine heatwaves Nature 604 486–90
Jacques-Dumas V, Ragone F, Borgnat P, Abry P and Bouchet F 2022 Deep learning-based extreme heatwave forecast Front. Clim. 4 1–14
Karevan Z and Suykens J A K 2020 Transductive LSTM for time-series prediction: an application to weather forecasting Neural Netw. 125 1–9
Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)
Laufkötter C, Zscheischler J and Frölicher T L 2020 High-impact marine heatwaves attributable to human-induced global warming Science 369 1621–5
Lguensat R, Sun M, Fablet R, Mason E, Tandeo P and Chen G 2018 EddyNet: a deep neural network for pixel-wise classification of oceanic eddies Int. Geoscience and Remote Sensing Symp. (IGARSS) vol 2018-July pp 1764–7
Milletari F, Navab N and Ahmadi S-A 2016 V-net: fully convolutional neural networks for volumetric medical image segmentation 2016 Fourth Int. Conf. on 3D Vision (3DV) pp 565–71
Mills K E et al 2013 Fisheries management in a changing climate: lessons from the 2012 ocean heat wave in the Northwest Atlantic Oceanography 26 191–5
Ñiquen M and Bouchon M 2004 Impact of El Niño events on pelagic fisheries in Peruvian waters Deep-Sea Res. II 51 563–74
Oliver E C J et al 2018 Longer and more frequent marine heatwaves over the past century Nat. Commun. 9 1–12
Oliver E C J et al 2019 Projected marine heatwaves in the 21st century and the potential for ecological impact Front. Mar. Sci. 6 1–12
Oliver E C J, Benthuysen J A, Bindoff N L, Hobday A J, Holbrook N J, Mundy C N and Perkins-Kirkpatrick S E 2017 The unprecedented 2015/16 Tasman Sea marine heatwave Nat. Commun. 8 1–12
Oliver E C J, Benthuysen J A, Darmaraki S, Donat M G, Hobday A J, Holbrook N J, Schlegel R W and Sen Gupta A 2021 Marine heatwaves Annu. Rev. Mar. Sci. 13 313–24
Qin J, Zhou L, Li B and Meng Z 2022 Prediction of the central Indian Ocean mode in S2S models Front. Mar. Sci. 9 1–10
Qin J, Zhou L, Li B and Murtugudde R 2020 Simulation of central Indian Ocean mode in S2S models J. Geophys. Res. Atmos. 125 1–17
Reynolds R W, Smith T M, Liu C, Chelton D B, Casey K S and Schlax M G 2007 Daily high-resolution-blended analyses for sea surface temperature J. Clim. 20 5473–96
Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th Int. Conf. (Munich, Germany) (Springer) pp 234–41
Sen Gupta A et al 2020 Drivers and impacts of the most extreme marine heatwaves events Sci. Rep. 10 1–15
Smale D A et al 2019 Marine heatwaves threaten global biodiversity and the provision of ecosystem services Nat. Clim. Change 9 306–12
Sun D, Jing Z, Li F and Wu L 2023 Characterizing global marine heatwaves under a spatio-temporal framework Prog. Oceanogr. 211 102947
Taylor J and Feng M 2022 A deep learning model for forecasting global monthly mean sea surface temperature anomalies Front. Clim. 4 932932
Thomson J A, Burkholder D A, Heithaus M R, Fourqurean J W, Fraser M W, Statton J and Kendrick G A 2015 Extreme temperatures, foundation species, and abrupt ecosystem change: an example from an iconic seagrass ecosystem Glob. Change Biol. 21 1463–74
Viglione G 2021 How heatwaves ravage the seas Nature 593 26–28
Vitart F 2017 Madden—Julian oscillation prediction and teleconnections in the S2S database Q. J. R. Meteorol. Soc. 143 2210–20
Vitart F et al 2017 The subseasonal to seasonal (S2S) prediction project database Bull. Am. Meteorol. Soc. 98 163–73
Wernberg T, Smale D A, Tuya F, Thomsen M S, Langlois T J, De Bettignies T, Bennett S and Rousseaux C S 2013 An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot Nat. Clim. Change 3 78–82
Wolff S, O'Donncha F and Chen B 2020 Statistical and machine learning ensemble modelling to forecast sea surface temperature J. Mar. Syst. 208 103347
Woo S, Park J, Lee J-Y and Kweon I S 2018 Cbam: convolutional block attention module Proc. European Conf. on Computer Vision (ECCV) pp 3–19