

Forecasting Marine Heatwaves using Convolutional Neural Networks

Jack Mei

Received October 20, 2024

Accepted May 23, 2025

Electronic access June 30, 2025

Marine heatwaves, caused by global warming, have been occurring more frequently than ever in recent years. Characterized by a period of unusually high sea surface temperatures, these natural events have severe implications for both the economy and the environment. This paper aims to create an algorithm which is able to predict marine heatwaves one month ahead and analyze which variables are crucial for the forecasting process. In order to achieve this, a convolutional neural network model was created with the U-Net architecture to encode and decode the information. The data was taken from the fifth generation ECMWF atmospheric reanalysis of the global climate (ERA5), separated into subgroups, and then analyzed across the ten year time period from 2011-2020. The results of each different subgroup were then analyzed to determine the strongest and weakest performing variables during the forecasting process. The model itself was also analyzed based on its accuracy when forecasting specific regions or features

1 Introduction

A marine heatwave is an event that occurs when a region in the ocean is experiencing unusually high temperatures. Although marine heatwaves occur naturally themselves, recent surges in global warming have amplified their effects and caused them to last significantly longer¹. Each marine heatwave can have disastrous impacts on the environment, through killing many unique species of fish, oysters, and coral, as well as causing the mass mortality of mammals and seabirds and harmful algal blooms². This in turn is a threat to global biodiversity, as effects of marine heatwaves extend to increasing coral bleaching, decreasing seagrass density, and decreasing giant kelp biomass³. Beyond just affecting the environment, marine heatwaves also damage the economy through tourism industries and fisheries, as the unfavorable conditions for aquatic species greatly reduces traffic for commercial and recreational fishing⁴.

Given the importance of marine heatwaves across the world, there has been substantial work to improve our understanding of the physical drivers of marine heatwaves and their predictability. Traditionally, physics based models and numerical methods have been used to forecast marine heat waves with some success. The performance of these traditional forecasting techniques are limited by our physical understanding of marine heatwaves, model resolution, and observational measurements. Specifically, traditional numerical models are run at too coarse a resolution to fully resolve all of the processes important to marine heatwaves. As a result, these sub-grid scale processes are parameterized, or estimated, often leading to poor performance and model bias. However, with the increase in satellite and in-situ observational datasets and recent developments in AI research, new methods such as logistic regression and decision trees have proven

to be a promising technique that can be used to improve our forecast skill of marine heatwaves⁵. One such model would be RandomForest, a decision tree-based algorithm, obtaining up to 76% accuracy when forecasting the formation and location of marine heatwaves weekly and 38% when forecasting their severity⁶. Furthermore, deep learning methods like recurrent neural networks and convolutional neural networks offer even more advanced capabilities compared to RandomForest, but require substantially more training data and are often computationally expensive to train. A Neural Basis Expansion Analysis for Interpretable Time Series (NBEATS) algorithm was found to have an accuracy of around 85% when predicting marine heatwaves 12 months in advance in the Gulf of Alaska, the Northeast Pacific, the West Coast of Australia, and the East China Sea⁷. These advancements highlight the potential for even greater improvements in forecasting accuracy, underscoring the need to investigate how other deep learning techniques can further refine predictions of marine heatwaves.

In this paper, we develop a convolutional neural network using the U-Net architecture to forecast global sea surface temperature one month in the future. We evaluate the significance of ten different variables for prediction and identify which ones resulted in the highest model accuracy.

The paper is organized as follows: in Section 2, we describe the methodology and machine learning pipeline we used to build the model. Section 3 reports the results of some notable models and compares them to other versions as well as the true sea surface temperature labels. The ultimate conclusions and outlook are summarized in Section 4.

#	Variable Names	Usage Styles
1	Sea Surface Temperature (SST)	Input, Output
2	Mean Sea Level Pressure (MSL)	Input
3	10 Meter Wind Speed (SI10)	Input
4	Surface Net Solar Radiation (SSR)	Input
5	Surface Net Thermal Radiation (STR)	Input
6	Total Precipitation (TP)	Input
7	Total Cloud Cover (TCC)	Input
8	Snow Albedo (SN)	Input
9	2 Meter Temperature (T2M)	Input
10	500 hPa Geopotential Pressure (Z)	Input

2 Methodology

2.1 Data Collection

The input data for this project was taken from the fifth generation ECMWF atmospheric reanalysis of the global climate (ERA5) dataset through the Copernicus Climate Data Store⁸¹. The measurements of sea surface temperature, mean sea level pressure (MSL), 10 meter wind speed (SI10), surface net solar radiation (SSR), surface net thermal radiation (STR), total precipitation (TP), total cloud cover (TCC), snow albedo (SN), and 2 meter temperature (T2M) were taken from the monthly averaged data on single levels, while the 500 hPa geopotential pressure (Z) was taken from the monthly averaged data on pressure levels. The variables were specifically chosen because of their connections to marine heatwaves, and each variable is on a 1 degree latitude and longitude grid between January 2011 to December 2020.

The variables were specifically chosen because of their connections to marine heatwaves. Since marine heatwaves are categorized by increases in sea surface temperature, this is the main indicating variable. Sea level pressure contributes to marine heatwaves by impacting the mixing of the ocean surface. With increases in pressure, there will be less mixing, therefore causing heat to build up. Similarly, a lower wind speed in certain regions will also decrease ocean surface mixing. Surface net solar and thermal radiation refers to the total amount of solar and longwave radiation being absorbed by the ocean surface after accounting for energy reflected by sources like clouds and the atmosphere. Cloud cover is a similar predictor because it controls the amount of direct sunlight reaching the ocean surface. Snow albedo data is mainly used to predict temperatures in colder areas like the Arctic, and areas with high concentrations of snow or ice are much more reflective of the solar radiation. Precipitation is an indicator for marine heatwaves since a higher ocean temperature results in an increase in evaporation. Geopotential pressure, which refers to the pressure at a certain altitude, can

trap warm or cold air over the ocean surface, causing changes in temperature and the formation of marine heatwaves. Each variable is on a 1 degree latitude and longitude grid between January 2011 to December 2020.

2.2 Data Loader

² The data loader was split into three methods: creating the custom dataset, loading and splitting the data, and saving the data loaders. During the creation of the custom dataset, the input variables are designated, and the data is retrieved at each index. Following this, the data is split into training, validation, and testing data loaders, with a split of approximately 80% (January 2011 - December 2018), 10% (January 2019 - December 2019), and 10% (January 2020 - December 2020), respectively. The model used the training data to adjust its learnable parameters, the validation data to check model accuracy, overfitting, and generalizability at each step, and the test data to calculate the final performance. During the preprocessing phase, each variable is normalized along the time dimension to ensure that each variable is non-dimensional and comparable using the equation

$$x = \frac{x - \mu}{\sigma}$$

where μ represents the mean and σ represents the standard deviation. In order to handle NaN values in the dataset, we replace them with values of 0. Since ERA5 is a global reanalysis product, it has perfect spatial and temporal coverage of the global oceans. As a result, the only regions where NaN values are swapped for 0 is over land, which is ensured to have no contribution to the loss or U-Net optimization because of the land-sea-mask applied before each optimization step. Finally, each data loader is saved to be later used in training and analysis.

2.3 U-Net Model

The convolutional neural network model used in this experiment was created using Pytorch with the UNet architecture using the traditional encode-decoder scheme, both with 3 computational blocks, and skip connections. Each encoder block consists of a convolution layer, a batch normalization layer, and a 2×2 pooling layer. The convolution layer converts the data into a matrix before multiplying it by a small 3×3 kernel matrix to transform it. Following this, the batch normalization layer is applied to normalize the data, which is then passed into the pooling layer that combines multiple outputs into a single neuron to reduce the dimensions of the data. After the convolutional layer is applied, the nonlinear relationships are captured using the rectified linear unit (ReLU) activation function. In each decoder block, there is a transposed convolution layer and another convolution layer to upsample the data back to its original dimensions.

While U-Nets are traditionally used for segmentation tasks, it is also appropriate for regression in forecasting sea surface tem-

1. <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means?tab=download>

2. <https://github.com/jack-meil/MHW-Research>

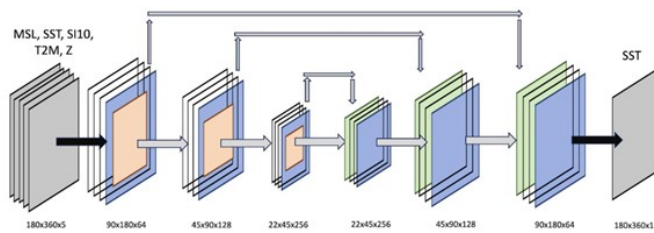


Fig. 1 This is an example of a 5-variable model taking in global, 1 degree latitude-longitude maps of mean sea level pressure, sea surface temperature, 10 meter wind speed, 2 meter temperature, and 500 hPa geopotential pressure. The color coding for the layers is as follows: white-colored tiles represent 3×3 convolution layers with ReLU activation, blue tiles represent batch normalization, orange tiles represent 2×2 max pooling layers, green tiles represent 2×2 transpose convolution layers, and the gray arrows represent concatenation and skip connection. The three numbers below each block represent the input height, width, and number of channels.

peratures. The encoder-decoder structure and skip connections allow the model to accurately address spatial relationships, preserving features like ocean eddies and upwelling zones through the upsampling and downsampling processes. However, one issue lies in the U-Net model’s ability to generalize and handle temporal patterns in marine heatwaves. This was slightly addressed by including variables at multiple timesteps, allowing the model to make predictions using recent trends. The inclusion of variables such as sea surface temperature, mean sea level pressure, and 2 meter temperature across all seasons also aids in capturing these patterns. Future works could integrate other architectures to better capture temporal patterns

In addition to the model itself, this category also contains a method to configure the optimizer. For this experiment, the AdamOptimizer algorithm was used to perform gradient descent. Adam Optimization is calculated with the formula

$$w_{t+1} = w_t - am_t$$

where w is the weight at time t , a is the learning rate which started with a value of 0.001, and m is the momentum at time t . The equation for the momentum is

$$m_t = m_{t-1} + (1 - \beta) \left[\frac{\delta L}{\delta w_t} \right]$$

with β representing the constant moving parameter defaulted to 0.9, δL representing the derivative of the loss function, and δw_t representing the derivative of the weight function at time t . AdamOptimizer was chosen over traditional gradient descent methods like Stochastic Gradient Descent (SGD) because it converges to the solution faster and reduces variance and noise during training.

2.4 Trainer

The purpose of the trainer class is to manage the data loading, the model weight optimization, and record the performance. Each model version was trained for 100 epochs; this number was chosen empirically to make sure the model was sufficiently accurate while stopping right before it starts to overfit.

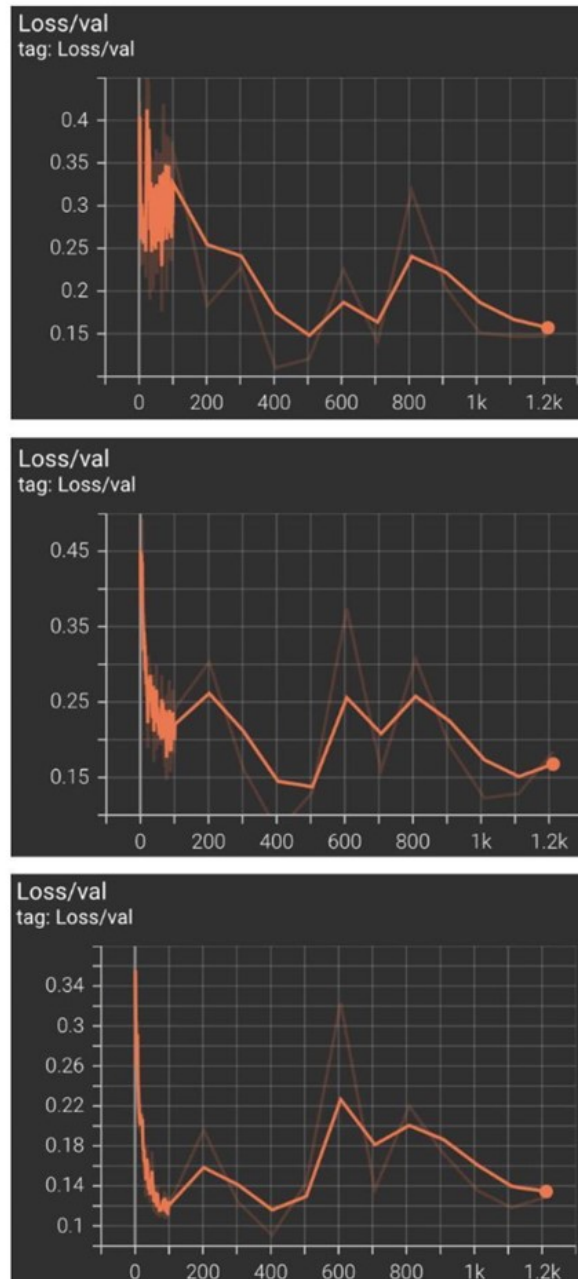


Fig. 2 These are the validation loss curves for the 10 variable, 5 variable v1, and 5 variable v2 models in descending order.

From the graphs shown, despite the variability in the valida-

tion loss as the steps increase, there is no clear trend of overfitting or underfitting. However, techniques like regularization can be used to improve the model in future studies. For each epoch, the gradients are set to zero so the parameter is updated without considering previous gradient accumulation, and the model is trained by processing batches of data of size 1. After each batch, the gradients for backpropagation are computed with gradient clipping to avoid exploding gradients and reduce noise in the loss curve. The running loss is printed at each step in the process, which is calculated using the Mean Squared Error (MSE) formula

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where y_i represents the output and \hat{y}_i represents the target value. The model is then set to evaluation mode, where the validation data is entered and loss is also calculated using MSE. At this point, a learning rate scheduler is used to ensure that the step sizes decrease as the loss approaches the absolute minimum value. The learning rate scheduler algorithm used in this experiment is Cosine Annealing, calculated through the formula

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_{\text{max}}} \pi \right) \right)$$

with η_{\max} being the initial learning rate, η_{\min} being the minimum learning rate, and T_{cur} being the current epoch. The model is saved every 20 epochs to minimize any losses due to a possible crash, and the loss is plotted on tensorboard to visualize the model's performance.

2.5 Analysis

The analysis module is the final step in the machine learning pipeline for this experiment. It takes in the most recent version of the model, usually after being trained for 100 epochs, and calls the saved test data loader. The test inputs are then fed into the model, and the model's outputs are compared to the test labels. The model evaluation is done after this process through plotting the test outputs on a world map and the test labels on another map, then comparing side by side or plotting another graph showing the difference between outputs and labels. Before plotting, the data had to be returned to its raw pre-normalized form by multiplying it by the standard deviation and adding the mean. The projections were accomplished using matplotlib, cartopy, and numpy. Cartopy is used for its PlateCarree world map outline, and every land mass is colored in gray since the main focus is on projecting sea surface temperatures. For the graphs created for side-by-side comparison, the color levels range from 265 Kelvin to 315 Kelvin in intervals of 5 Kelvin,

and these values were customized to make sure the colors were sufficiently visible and distinct. When graphing the differences between model predictions, a different color range from - 20 Kelvin to 22 Kelvin with intervals of 2 Kelvin was used due to the reduction in color intensity. These differences are then quantified using the Pearson correlation coefficient equation, or

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where x represents the predicted values, y represents the true values, $\sum(x_i - \bar{x})(y_i - \bar{y})$ represents the covariance between x and y , $\sum(x_i - \bar{x})^2$ represents the standard deviation for x , and $\sum(y_i - \bar{y})^2$ represents the standard deviation for y . The square of the r value, r^2 , represents the explained variance, and is calculated by vectorizing or flattening the prediction and the truth maps. This quantifies the similarity between the two maps and is known as pattern correlation, which can then be used to analyze the time and locations where the model is performing either poorly or well. The closer the r^2 value is to 1, the more accurate a graph is.

3 Results

3.1 Ten Variable Model

The first model for examination included every single tested variable, including sea surface temperature, mean sea level pressure, 10 meter wind speed, 2 meter temperature, snow albedo, surface net solar radiation, surface net thermal radiation, total cloud cover, total precipitation, and 500 hPa geopotential pressure. For select months, the model predictions were as follows: (Figure 3)

At just a glance, the model is able to qualitatively forecast the large scale sea surface temperature patterns, such as larger heat concentrations towards Antarctica and colder sea surface temperatures off the coast of Malaysia. The model also captures cooling on the West coast of the United States, which is likely associated with the North Pacific gyre and coastal upwelling. The North Pacific Gyre pulls deep cold water and nutrients to the surface, and upwelling is typically strongest in the winter, which is reflected in the prediction. The warmer areas between South America and Africa are also being somewhat properly captured, alongside a large patch of higher temperatures in the Indian Ocean. However, the model performs poorly when forecasting the intensity of sea surface temperatures associated with individual ocean eddies. The explained variance r^2 value of 0.863 further reinforces the accuracy of the model at predicting global sea surface temperatures at a monthly lead time.

Similar patterns can be noticed between Figure 4 and Figure 3, where widespread warm sea surface temperature north of Russia can be noticed in observations but not in our forecast. From the difference plot, it is shown that the model accurately

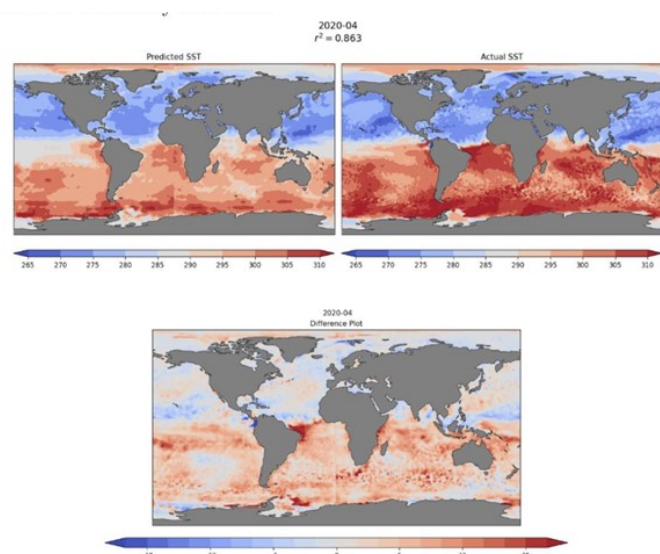


Fig. 3 is graph shows a side-by-side comparison between the 10-variable prediction and the real sea surface temperatures in April, when it's winter in the northern hemisphere and summer in the southern hemisphere. At the bottom, a graph showing the differences between the actual and predicted sea surface temperatures was generated by subtracting the predicted values from the labels.

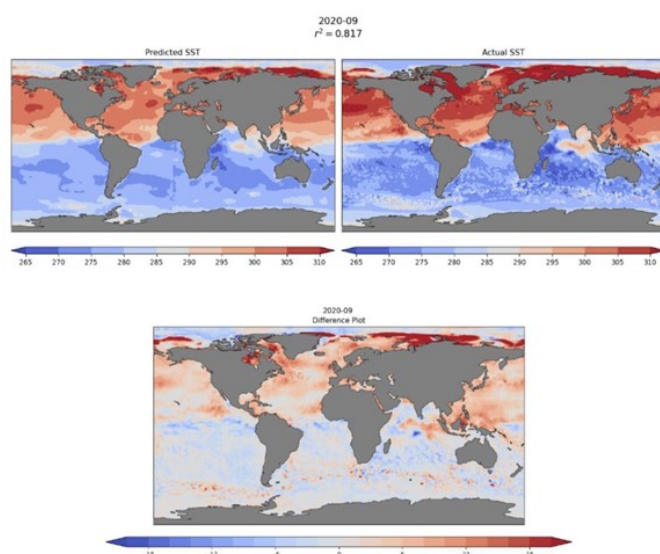


Fig. 4 This graph shows the side-by-side comparison in September, when it is summer in the northern hemisphere and winter in the southern hemisphere.

predicted most of the southern hemisphere winter temperatures, excluding a few small eddies, while forecast skill near the North Pole appears to be substantially worse. We hypothesize that this reduced forecast skill at high latitudes may be attributable to the lack of information about sea ice as input to our model. There is also an indication of colder temperatures off of the east coast of Africa, spreading throughout parts of the Indian Ocean. The model accurately captures that the coldest temperatures were closest to the east coast of Somalia and Kenya, and the intensity gradually decreases as distance increases. The r^2 value of 0.817, though not as high as shown in Figure 3, is high enough to show that the model is correctly learning to identify the major processes governing the evolution of sea surface temperature on monthly timescales.

3.2 Five Variable Model Version 1

The five chosen variables in this version were sea surface temperature (SST), mean sea level pressure (MSL), 500 hPa geopotential pressure (Z), 2 meter temperature (T2M), and 10 meter wind speed (SI10).

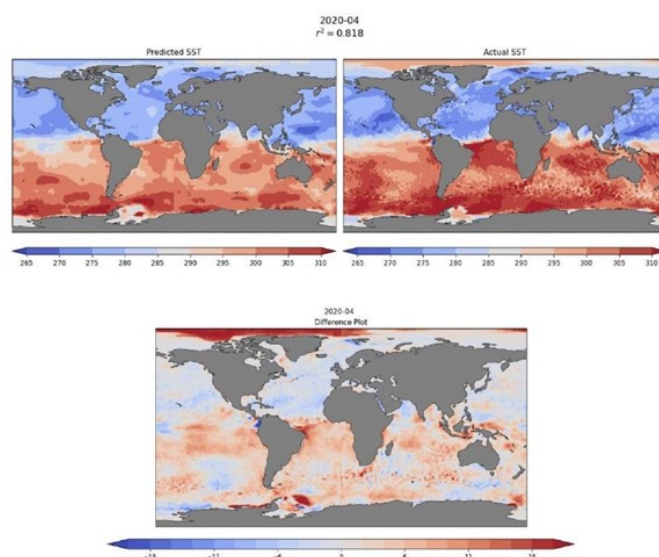


Fig. 5 This shows the results of a 5-variable model in April of 2020 using sea surface temperatures, mean sea level pressure, 500 hPa geopotential pressure, 2 meter temperature, and 10 meter wind speed.

Once again, the main inaccuracies that are highlighted by the truth graph include differences in sea surface temperature intensity and the lack of attention towards the small ocean eddies. The North Pacific Gyre is still somewhat predicted off the West coast of the United States, and there are some significant discrepancies off the coast of Brazil, next to the Weddell Sea in Antarctica, and around the North Pole. Encouragingly, the model clearly captures the seasonal hemispheric asymmetry of sea surface temperatures with the Northern hemisphere being

substantially cooler than the Southern hemisphere.

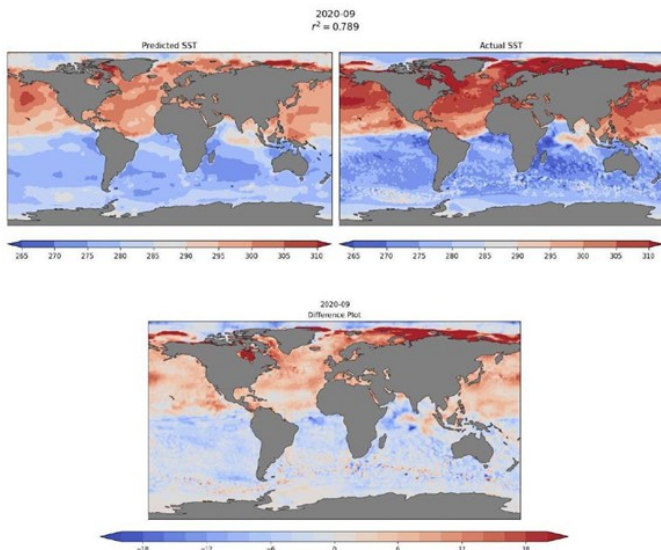


Fig. 6 This shows the results of the same model in Fig 4 in September of 2020.

In September, with summer in the northern hemisphere and winter in the southern hemisphere, the patterns of prediction change slightly. There is no longer a large discrepancy in the Weddell Sea; however, the particularly high sea surface temperatures near the North Pole still remains undetected. Another notable problem is that there is still trouble predicting temperature intensity, especially in the warmer summer regions. Despite this, the model is still performing well enough to capture almost the entire trend in sea surface temperatures off of the East coast of Africa.

3.3 Five Variable Model Version 2

The second version of the five variable model contained values for sea surface temperature (SST), surface net solar radiation (SSR), surface net thermal radiation (STR), 10 meter wind speed (SI10), and 500 hPa geopotential pressure (Z). Compared to the first version of the 5-variable model, data for the mean sea level pressure and 2 m temperature were replaced with both surface net thermal and solar radiation. Solar radiation, or shortwave radiation, is a measure of how much sunlight is being absorbed, while thermal radiation, or longwave radiation, refers to black-body radiation from sources like greenhouse gasses, water vapor, and clouds. The calculated net radiation is the total amount of emitted solar and thermal radiation subtracted from the total amount of radiation absorbed.

This model had more mixed results compared to its previous counterparts. While the other versions stayed consistent in their mispredicted temperatures and intensities, the difference plot shows how this model gave more random results, with the areas

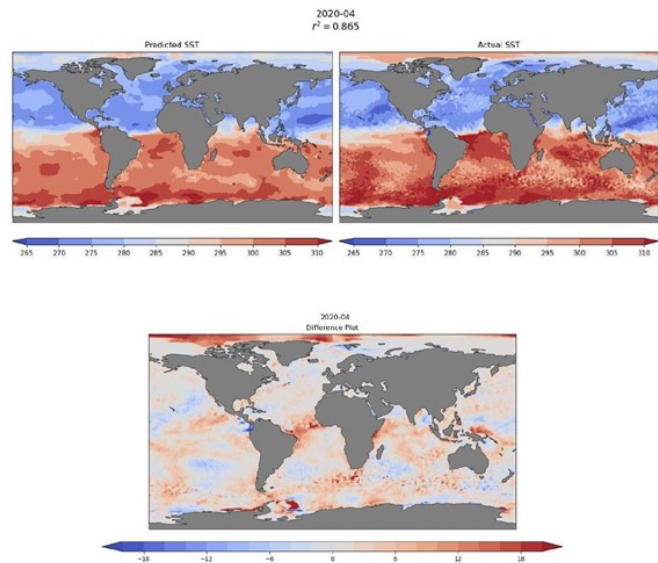


Fig. 7 This shows the results of a 5-variable model in April of 2020 using sea surface temperatures, 10 meter wind speed, 500 hPa geopotential pressure, and surface net solar and thermal radiation.

predicted to be hotter or colder than the labels seeming much more sporadic in the southern hemisphere. Apart from that, the high temperatures in the North Pole, off the coast of Brazil, and in the Weddell Sea remain difficult to predict. There is also a slight error off the coast of Indonesia and Papua New Guinea where the model predicted a cooler temperature than the label. However, the model performs well around South America, as the high temperatures off of the west coast of Brazil were accurately captured, although the intensity is a bit off. The slightly cold portion to the west of Colombia was also captured well along with the hotter areas around it. Switching seasons to summer in the northern hemisphere and winter in the southern hemisphere, the model produced the following results: (Figure 8)

Compared to the April predictions, the Figure 8 predictions of high temperatures in the North Pole as well as the Weddell Sea region are significantly more accurate. As mentioned above, this is likely due to the impact of sea ice concentration on sea surface temperatures, since this model is not provided with any sea ice data. However, one noticeable difference is that the spot off the coast of Papua New Guinea and Indonesia seems to be a bit harder to predict, as the results are worse compared to earlier in the year. Still, the model properly captured the colder temperatures in the area between Mozambique and Madagascar as well as on the east of Madagascar. The prediction of high temperatures above Russia are also covered relatively well.

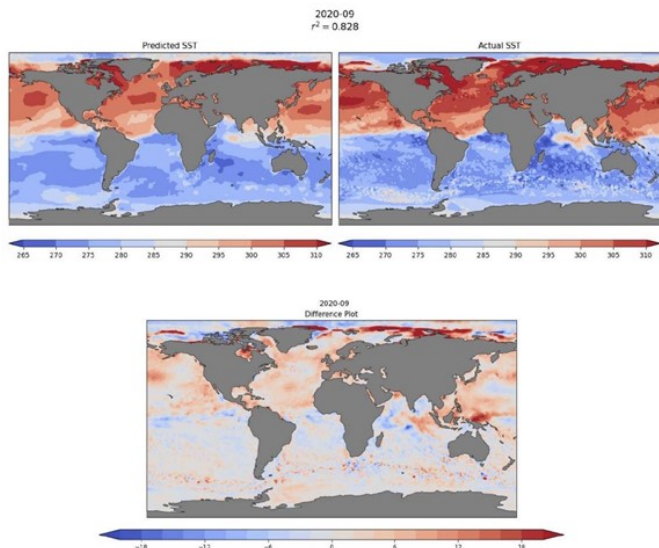


Fig. 8 This shows the results of the same model in Fig 6 in September of 2020.

3.4 Comparisons

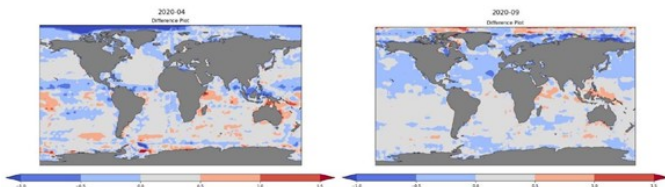


Fig. 9 These graphs show the differences between the 10-variable model and the first version of the 5-variable model taking in mean sea level pressure and 2 meter temperature in both April and September.

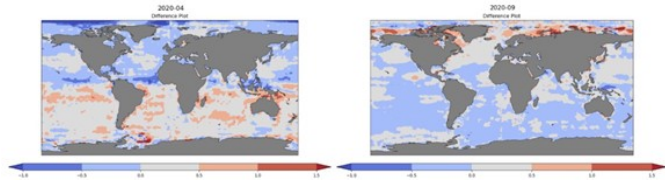


Fig. 10 These graphs show the differences between the 10-variable model and the second version of the 5-variable model taking in surface net solar and thermal radiation in both April and September.

From the results shown, the second version of the 5 variable model performed much better than the first version in both April and September, which demonstrates that surface net solar and thermal radiation as indicators improve model performance more than mean sea level pressure and 2 meter temperature.

Mean sea level pressure leads to an increase in sea surface temperatures because of its effect on ocean circulation. High pressure zones cause downwelling, which reduces the mixing of cold water from deeper areas, resulting in a higher sea sur-

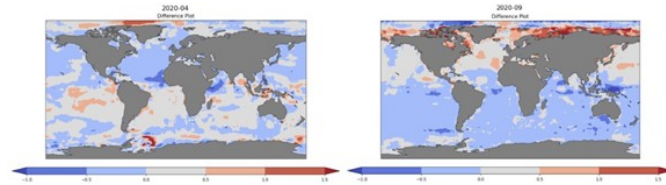


Fig. 11 These graphs show the differences between the two different 5-variable models, the first taking in mean sea level pressure and 2 meter temperature and the second taking in surface net solar and thermal radiation, in both April and September.

face temperature. Low pressure zones are the opposite, with upwelling decreasing the sea surface temperature. 2 meter temperature has similar impacts on ocean mixing, with high temperatures corresponding to more stable atmospheric conditions and less mixing while lower temperatures correspond to strong winds and more mixing. On the other hand, surface net solar and thermal radiation have a more direct impact on the changes in sea surface temperature. Surface net solar radiation refers to the energy from the sun reaching the ocean, which decreases as cloud cover increases to reflect the energy. Surface net thermal radiation is energy being emitted back into the atmosphere by the ocean, which increases when there are no clouds to trap the heat. As a result, a higher surface net solar radiation and a lower surface net thermal radiation directly correlate with an increase in sea surface temperature.

Overall, the predictions of the 10-variable and 5-variable models differed by around 1-2 degrees Kelvin, mainly only exceeding that threshold up in the North Pole in every single month. This is most likely because the model does not take sea ice data into consideration for the prediction, and melting ice caps could either cause the sea surface temperatures to increase or decrease drastically. Another observation is that the difference between the 10-variable and 5-variable models is much less noticeable compared to the difference between the 5-variable models. A possible explanation for this is the diminishing returns when adding in variables, since adding multiple closely related variables to the dataset could decrease the model's accuracy. For example, including both snow albedo and net solar radiation may possibly have decreased the performance due to snow albedo's effects on the variables in the mixed layer heat budget equation for radiation. By comparing the two 5-variable models to each other, it becomes easier to identify how each model impacts the forecasts because better forecasted regions are more controlled by net solar and thermal radiation.

	10 VAR	5 VAR V1	5 VAR V2
2020 - 04	0.863	0.818	0.865
2020 - 09	0.817	0.789	0.828

Fig.12 The performance of each model in each month as measured by the r^2 value is shown here for comparison purposes.

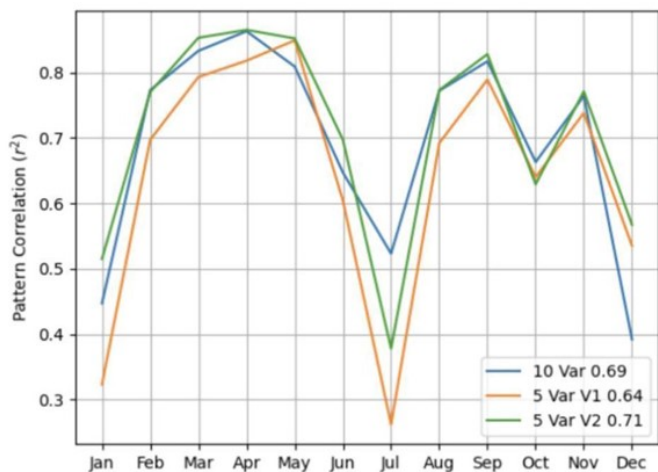


Fig 13. This graph shows the pattern correlation values for explained variance compared across all three models to analyze performance.

The model appears to perform far better during the "shoulder" or transition seasons (i.e. fall and spring), and worse during the peak seasons (summer and winter). These results suggest that the model performs best when the hemisphere temperature asymmetry is weakest. This may be because we do not provide the model with information about atmospheric heat transport or top of atmosphere solar insolation (i.e. the amount of sunlight during that time of year). However, this discrepancy warrants further investigation and will be the subject of future study.

3.5 Discussion

The findings from this research on seasonal variations in model accuracy and limitations in forecasting sea surface temperatures have important implications for understanding climate patterns and improving predictive models. Most notably, there are limitations with regards to using convolutional neural networks for map projections in general, as the downsampling and up-sampling process smoothes out all of the data, which is one of the many reasons why the model struggles to predict small ocean eddies. Additionally, the noticeable line in the middle of each prediction map is caused by discontinuities since the latitude-longitude map being used is acyclic, but our globe is spherical. A more ideal version to solve this issue would be through implementing custom padding by taking the left and right edges and concatenating them onto each other.

Due to the training and technology limitations for this paper, the model was only able to be trained for a maximum of 10-variables forecasting 1 month ahead, and other more sophisticated techniques are out of the scope of this paper. However, further studies to expand on these results include but are not limited to plotting performance error and adding explainable

AI techniques. Specifically, a study on Shap to assign a value to each variable to show how it contributed to the model's predictions would be extremely helpful to figure out which data points were the most useful and which did not have much of an impact⁹. This could also help reduce the diminishing returns when adding in closely related variables, which would in turn increase the overall forecasting accuracy.

4 Conclusion

In conclusion, this study successfully developed a convolutional neural network model utilizing the UNet architecture to predict marine heatwaves one month in advance, addressing a critical need in climate science and resource management. By leveraging data from the ERA5 atmospheric reanalysis, various environmental variables were identified and analyzed, revealing significant insights into their impact on forecasting accuracy. The findings highlight the strongest predictors of marine heatwaves, which can inform future research and operational forecasting efforts. Additionally, the model's performance demonstrated variability across different regions, suggesting the necessity for localized approaches in marine heatwave predictions. As marine heatwaves continue to pose substantial risks to ecosystems and economies, this research not only contributes to the understanding of these phenomena but also provides a foundational tool for more effective climate adaptation strategies. Future work should focus on refining the model and exploring additional datasets to enhance its robustness and applicability across diverse marine environments. Despite the shortcomings discussed above, this paper demonstrates that AI can be used for marine heatwave forecasting and encourages others in the scientific community to participate in advancing this vital area of research.

References

- 1 T. L. Frlicher, E. M. Fischer and N. Gruber, *Nature*, 2018, **560**, 360–364.
- 2 K. E. Smith, M. T. Burrows, A. J. Hobday, N. G. King, P. J. Moore, A. S. Gupta, M. S. Thomsen, T. Wernberg and D. A. Smale, *Annual Review of Marine Science*, 2022, **15**, 119–145.
- 3 D. A. Smale, T. Wernberg, E. C. J. Oliver, M. Thomsen, B. P. Harvey, S. C. Straub, M. T. Burrows, L. V. Alexander, J. A. Benthuisen, M. G. Donat, M. Feng, A. J. Hobday, N. J. Holbrook, S. E. Perkins-Kirkpatrick, H. A. Scannell, A. S. Gupta, B. L. Payne and P. J. Moore, *Nature Climate Change*, 2019, **9**, 306–312.
- 4 K. E. Smith, M. T. Burrows, A. J. Hobday, A. S. Gupta, P. J. Moore, M. Thomsen, T. Wernberg and D. A. Smale, *Science*, 2021, **374**, year.
- 5 G. Bonino, G. Galimberti, S. Masina, R. McAdam and E. Clementi, *Ocean Science*, 2024, **20**, 417–432.
- 6 K. Giamalaki, C. Beaulieu and J. X. Prochaska, *Geophysical Research Letters*, 2022, **49**, year.
- 7 A. Prasad, S. Sharma and H. Agarwal, *Forecasting Marine Heatwaves using Machine Learning*, EarthArXiv (California Digital Library), 2022.

-
- 8 H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Hornyi, J. Muoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. D. Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hlm, M. Janiskov, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. Rosnay, I. Rozum, F. Vamborg, S. Villaume and J. Thpaut, *Quarterly Journal of the Royal Meteorological Society*, 2020, **146**, year.
- 9 B. Rozemberczki, L. Watson, P. Bayer, H. Yang, O. Kiss, S. Nilsson and R. Sarkar, *The shapley value in machine learning*, arXiv preprint arXiv:2202.05594, 2022.