

## OCEANOGRAPHY

# Purely satellite data–driven deep learning forecast of complicated tropical instability waves

Gang Zheng<sup>1</sup>, Xiaofeng Li<sup>2,3\*</sup>, Rong-Hua Zhang<sup>2,3</sup>, Bin Liu<sup>4</sup>

Forecasting fields of oceanic phenomena has long been dependent on physical equation–based numerical models. The challenge is that many natural processes need to be considered for understanding complicated phenomena. In contrast, rules of the processes are already embedded in the time-series observation itself. Thus, inspired by largely available satellite remote sensing data and the advance of deep learning technology, we developed a purely satellite data–driven deep learning model for forecasting the sea surface temperature evolution associated with a typical phenomenon: a tropical instability wave. During the testing period of 9 years (2010–2019), our model accurately and efficiently forecasts the sea surface temperature field. This study demonstrates the strong potential of the satellite data–driven deep learning model as an alternative to traditional numerical models for forecasting oceanic phenomena.

## INTRODUCTION

Deluges of satellite-derived ocean products not only provide an unprecedented golden opportunity for in-depth research but also demonstrate the urgent need to develop useful methods to explore time-series observation. Sea surface temperature (SST) is one of the ocean products that has the most extended history and a critical parameter to help people understand scientific questions in physical/biological oceanography and atmosphere-ocean interaction. Since becoming relatively easy to measure from space with high accuracy, SST has been widely used to reveal various critical oceanic phenomena, e.g., tropical instability wave (TIW) (1). Traditional statistical analyses in previous studies have limitations in model complexity in addressing events that are complicated by nature.

As one of the most popular and influential deep learning (DL) techniques and the improved version of an artificial neural network, the deep neural network (DNN) technique adopts deep network architectures and efficient weight-shared convolutional layers (2, 3). This technique allows a DNN-based DL model to be substantially more complex than traditional statistical models. It can more effectively mine complicated rules deeply hidden in a time series of multi-dimensional observation data by training on a large amount of known sample data. The applications of DL models in oceanography and other fields of geosciences are still in their infancy (4–6). The similarities between SST pattern forecasting and video frame prediction inspire us to believe that DL technology offers promise for building new data-driven models to study oceanic phenomena.

Recently, considerable attention has focused on using both site-specific and site-independent artificial neural network models to forecast SSTs. A site-specific model uses separate artificial neural network models to predict SSTs at different sites (7). The model considers the site difference but has high computational costs and requires sufficient training data at each location. In contrast, the more efficient site-independent models use the same model for all

sites (8, 9). Two oceanic phenomena represented by the SST patterns could still be related to each other even if these two phenomena are at two relatively remote sites. For example, the SST patterns at two distant sites could still always be associated with TIWs that have large wavelengths of hundreds to thousands of kilometers (10–16). However, all these recent models forecast a future SST at one site only based on the prior SST series at this forecast site or its very close neighboring sites without considering the information hidden in the SST series at the other remote locations (7–9).

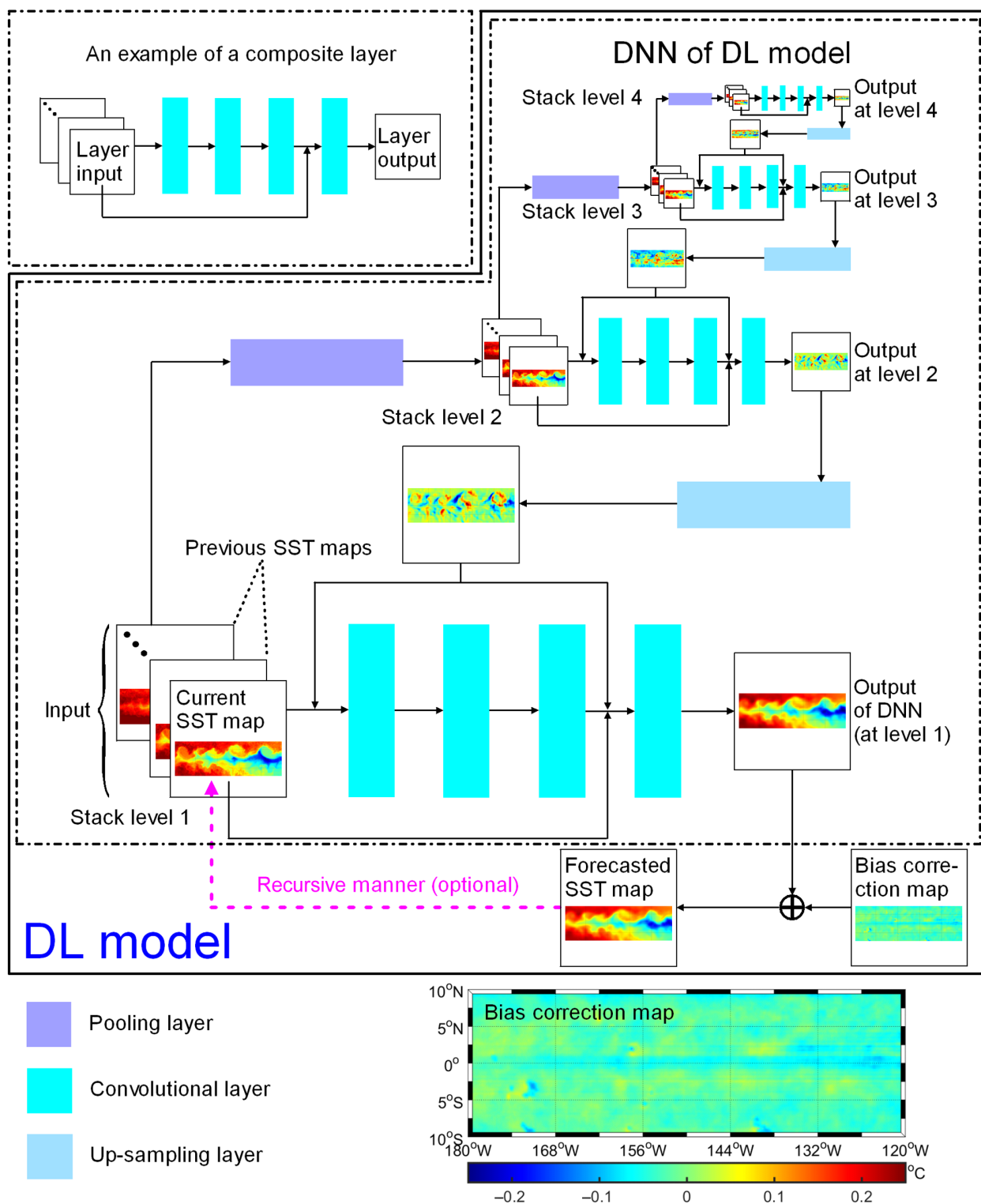
Pacific TIW, a prominent prevailing oceanic phenomenon in the eastern equatorial Pacific Ocean, is characterized as cusp-shaped waves propagating westward at both flanks of the equatorial Pacific cold tongue. Estimations for the TIW wavelengths, periods, and phase speeds have been generated from various data sources, therein having typical values of 600 to 2000 km, 15 to 40 days, and 17 to 86 km/day (10–16). Forecasting TIWs is challenging because of the high complexity. TIWs affect heat, mass, and momentum transport in the ocean dynamics (15–19), air-sea (20) and biophysical (21–23) interactions, climate change (24–26), among others, and, conversely, are also affected by these processes in turn. Using physical equation–based numerical modeling to study TIWs, one needs to consider complicated processes to make the models more realistic, which is a difficult task.

In this study, we mainly developed a purely data-driven model to predict SST variation. Our data-driven model is based on the DL technology and named the DL model for brevity. The DL model consists of a DNN and a bias correction map (see Fig. 1). The DL model can automatically learn rules of SST pattern spatial-temporal variations associated with TIWs from satellite data and avoids explicitly modeling various complicated processes with physical equations. The DL model was tested for 9 years that did not overlap with the training period.

In the literature, people also developed several statistical and artificial intelligence models for forecasting different ocean parameters: waves, precipitation, polar sea ice, sea surface height, SST, etc. (7–9, 27–31). We compared our DL results against three popular ones, including the site-specific linear regression model, the multi-layer perceptron model, and the long short-term memory model. The site-specific linear regression model is very efficient for modeling linear relationships with a simple form and was independently

Copyright © 2020  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China. <sup>2</sup>CAS Key Laboratory of Ocean Circulation and Waves, Big Data Center, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. <sup>3</sup>Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China. <sup>4</sup>College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China. \*Corresponding author. Email: lixf@qdio.ac.cn



**Fig. 1. Architecture of our DL SST forecast model.** The model consists of a DNN and a time-independent bias correction map, both of which were determined by the samples during the training period. The model receives SST maps at the previous and current time steps and then outputs the SST map at the future time step. The DNN has four stacked composite layers, each of which receives SST maps at different resolutions and has four cascaded convolutional layers. The bias correction map is added to the DNN output to obtain the forecasted SST map.

applied to different forecast sites here. The multilayer perceptron model, inspired by biological neural networks, has a robust fitting capability when relationships are nonlinear. The long short-term memory model is a recently developed type of recurrent neural network and has become one of the most popular networks for sequence modeling. Compared with the three models (the detailed comparisons are given in the Supplementary Materials), the DL has better performance.

We used the DL model to describe TIWs. As will be shown later, the DL model successfully made the short-term (5-day) forecast of the SST field associated with TIWs, and captured the propagation of TIWs. Our study shows the strong potential of DL technology for forecasting fields of complicated oceanic phenomena.

## RESULTS

### Data for training and testing the DL model

The blended daily global SST version 5.0 products from Remote Sensing Systems were used to build and test our DL model. These gridded SST products are made from measurements by both space-borne microwave sensors (of Tropical Rainfall Measuring Mission Microwave Imager, Advanced Microwave Scanning Radiometer-E/-2, WindSat Polarimetric Radiometer, and Global Precipitation Measurement Microwave Imager) and infrared sensors (of Moderate Resolution Imaging Spectroradiometer-Terra/-Aqua and VIIRS-NPP). We extracted data in the eastern equatorial Pacific Ocean from 10°S to 10°N and 180°W to 120°W. The resolution is 9 km × 9 km; however, we reduced it to 18 km × 18 km by spatial box averaging, and the total number of grids is 116 × 348. The SST data span 13 years between 2006 and 2019, and we divided this period into two non-overlapping periods from 1 January 2006 to 31 December 2009 and from 1 January 2010 to 31 March 2019. The first 4 years of data were used to train our DL model. The second 9 years of data were used to test the performance of our DL model. Because TIWs are internally generated ocean waves with time scales of approximately 15 to 40 days, we set our DL model time step to 5 days. The time series of the SST maps at the previous 13 and current time steps (i.e., the 1st to 14th steps) were used to forecast the SST map at the next future time step (the 15th step). Therefore, the data dimension is 15. We then shifted the time series 1 day at a time to build the second, third, fourth, etc., SST map series. In each time series, the SST maps at the first 14 steps were input to the DL model, and the SST map at the 15th step was used as the ground truth to validate the DL model forecast. Collectively, we obtained 1391 and 3307 samples of SST map series during the training and testing periods, respectively.

### Our data-driven DL SST forecast model

As shown in Fig. 1, the DL model consists of a trained multiscale DNN and a time-independent bias correction map. We make the next-time-step SST forecast using satellite SST maps at the previous 13 and current time steps. For example, the forecast on 12 September 2010 is made using 14 satellite SST maps at a 5-day interval between 4 July ... 2 September and 7 September. The DNN has four stacked composite layers, each of which processes the input SST map series at different spatial resolutions. We designed this multiscale network architecture, considering that the SST patterns are related to each other within a vast area of the ocean. The idea of a multiscale scheme has achieved notable successes in the field of computer vision, e.g., DNN applications in semantic segmentation (32). We also consid-

ered the natural differences among different sites, so we further corrected the DNN forecast by adding the bias correction map to get the DL-forecasted SST map. The DL model was trained and tested in two nonoverlapping periods of 4 years (2006 to 2009) and 9 years (2010 to 2019). More details on the DL model development are given in Methods.

### Accuracy during the testing period

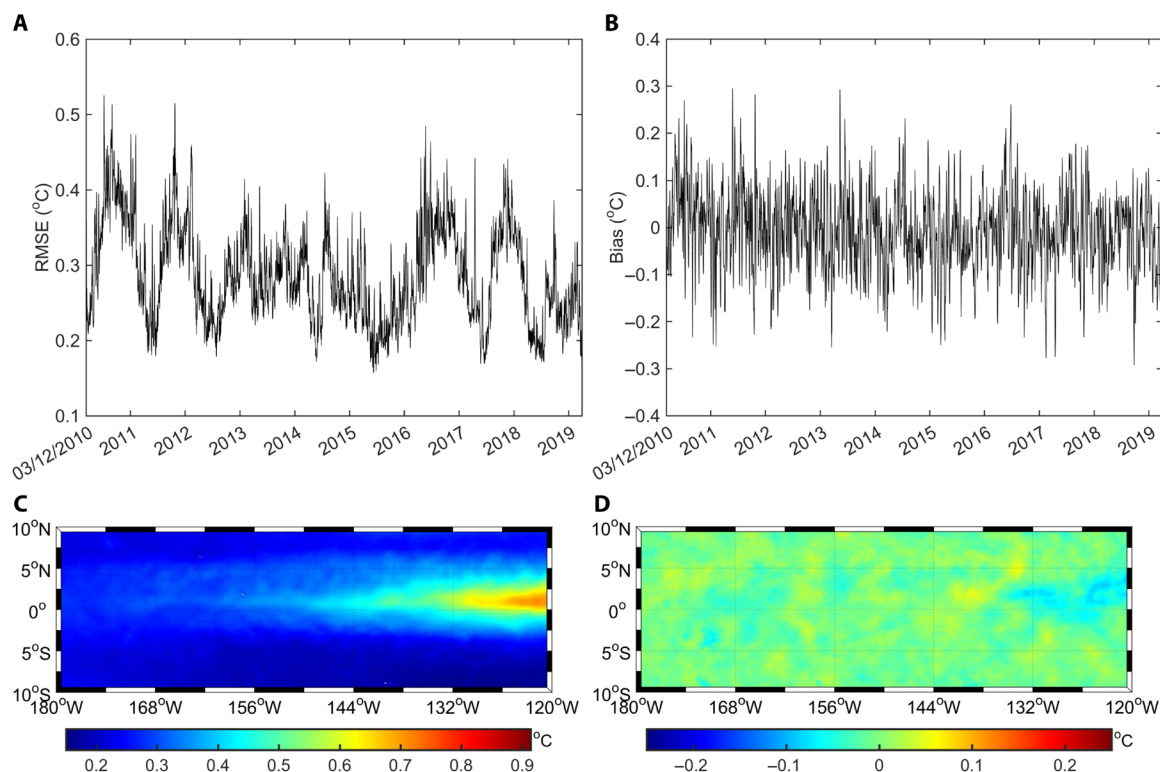
The temporal trends of root mean square error (RMSE) and bias of the DL model were calculated sample by sample from the forecasting errors at all grids during the testing period of 9 years and presented in Fig. 2 (A and B). In general, the RMSE and bias are stable, fluctuating between 0.15° to 0.45°C and −0.15° to 0.15°C, respectively. As will be shown later, the RMSE peaks occur during the TIW seasons caused by the TIW-induced fast SST pattern variation. The RMSE and bias spatial distributions of the DL model were calculated grid by grid from the forecasting errors of all samples during the testing period and are given in Fig. 2 (C and D). The RMSE is higher in the cold tongue area than in the other region. This high RMSE is due to the large SST spatial gradient and fast temporal variation of the cold tongue. The global RMSE and bias of all grids and all samples in the study area during the testing period are 0.29° and −0.01°C, respectively.

### SST pattern forecast of TIW motion

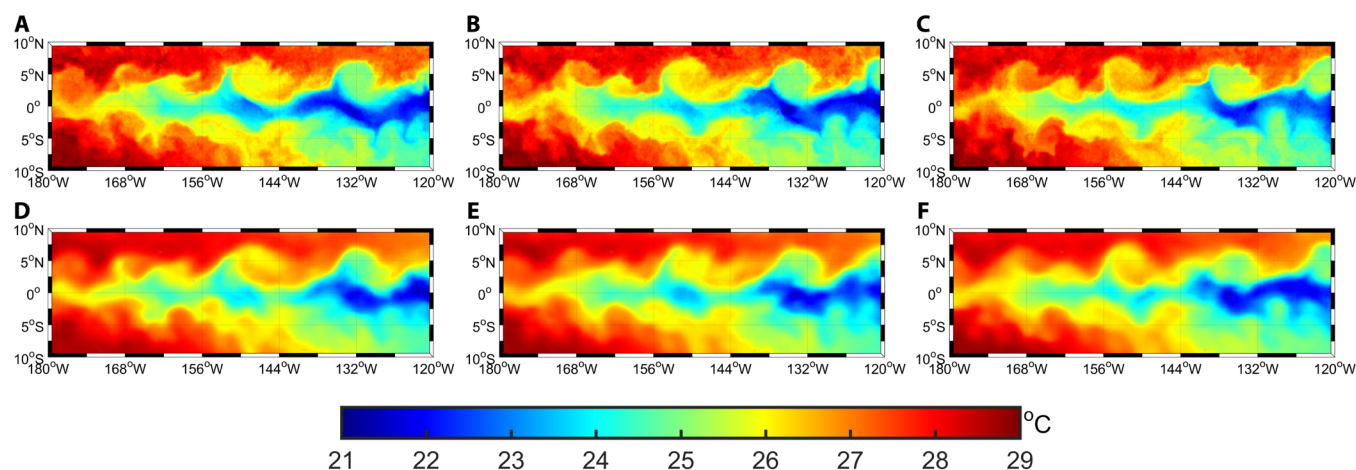
Figure 3 (A to F) shows the satellite and forecasted SST maps at three consecutive time steps during the testing period (see movie S1 for satellite and forecasted SST maps at more time steps). One of the most prominent SST patterns along the equatorial ocean is the cusp-shaped characteristic of TIWs that propagate westward with irregular deformations. The SST maps match closely in shape.

The outputs at the three consecutive time steps of the four composite layers in Fig. 1 are visualized from the first (bottom) to the fourth (top) stack level of the DNN (fig. S1). For clarity, the coarser resolutions at higher levels are converted to the original resolution by nearest-neighbor interpolation, and the outputs are linearly normalized to [−1, 1] by using the minimum and maximum values of the outputs as two ends. All the outputs present the maps visually relevant to the satellite SST maps in Fig. 3 (A to C) and show the signal of westward propagation. These maps are the TIW spatial-temporal features extracted by the DNN with the rules automatically learned from the samples during the training period of 4 years. The DNN used the features to forecast the motion of the TIWs.

The meridional averages (MAs) of the satellite and forecasted SST maps were calculated. We estimated the speed of the SST pattern westward propagation on the basis of the maximum detrended cross-correlation along the equator between the MAs at the current time step and the next time step (see Methods). Figure 4A shows the estimated speeds within a normal speed range of 0 to 100 km/day (10–16). The forecasted SST pattern propagation speed (green solid curve) is in good agreement with that (deep brown dashed curve) estimated from the satellite/satellite SST MA pairs. TIWs are generated by the barotropic and baroclinic instability processes due to the meridional and vertical shear of the equatorial currents (12, 33, 34). Consequently, TIWs are active/inactive during boreal fall/spring, as the current shear is stronger/weaker. During TIW seasons, the motion of the SST pattern is dominated by the TIWs. The forecasted SST pattern propagation speed can be considered as the TIW speed. However, during the no- or weak-TIW seasons, the SST pattern is inert, with no apparent westward motion. The RMSE of the DL



**Fig. 2. Accuracy of DL model during the testing period.** (A) Temporal trend of RMSE, (B) temporal trend of bias, (C) spatial distribution of RMSE, and (D) spatial distribution of bias. The RMSE and bias temporal trends were calculated sample by sample from the forecasting errors at all grids. The RMSE and bias spatial distributions were calculated grid by grid from the forecasting errors of all samples.



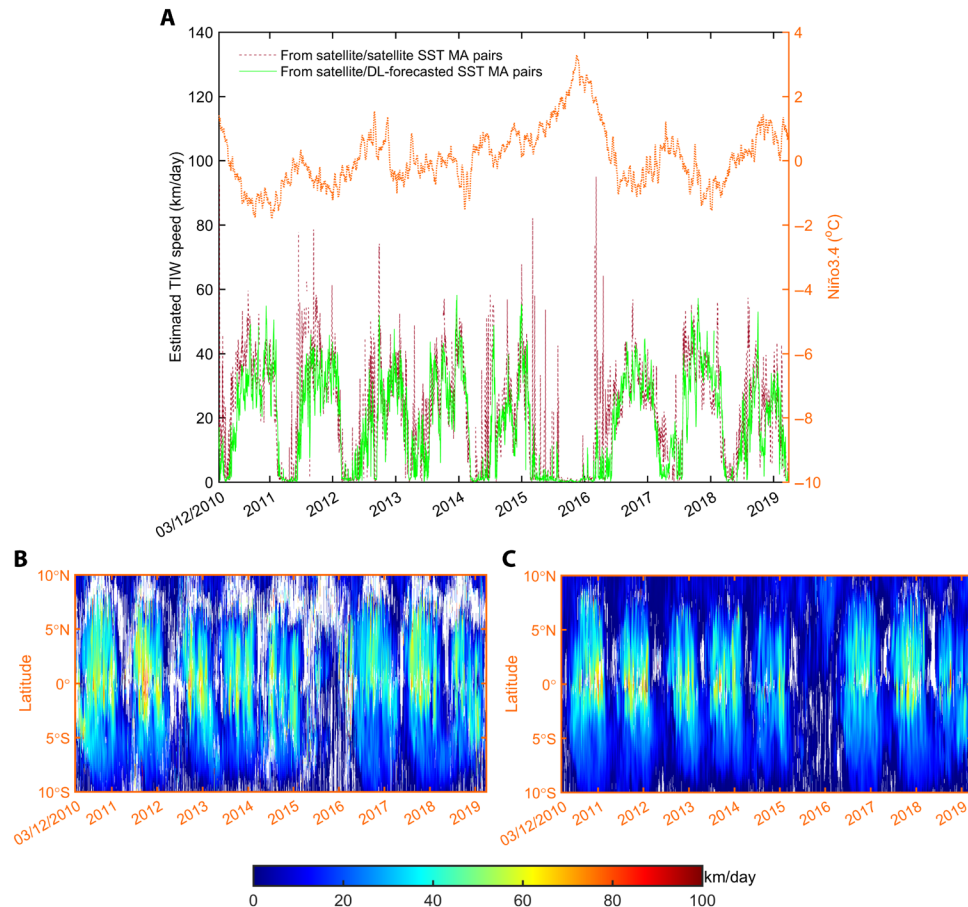
**Fig. 3. Satellite and DL-forecasted SST maps at three consecutive time steps (12, 17, and 22 September 2010) with an interval of 5 days.** We make the next-time-step SST forecast using satellite SST maps at the previous 13 and current time steps. For example, the forecast on 12 September 2010 is made using 14 satellite SST maps at a 5-day interval between 4 July ... 2 September and 7 September. (A to C) Satellite SST maps at the three time steps and those forecasted by the DL model (D to F).

model is more significant during the TIW seasons (Fig. 2A) due to the fast SST pattern variation.

The daily Niño3.4 index data provided by the KNMI (the Royal Netherlands Meteorological Institute) Climate Explorer are also overlaid in Fig. 4A (orange dotted curve). We can see that the Niño3.4 index values and the forecasted TIW speed values are 180° out of phase. During the major El Niño event from 2014 to 2016, the

TIW speed values were almost zero due to the weakening of meridional SST gradients. This fact is also consistent with the mooring and Argo float measurements from 2000 to 2010, where TIW kinetic energy and occurrence probability were negatively correlated with the Niño3.4 index (18). After removing the speed outliers beyond the normal range of 0 to 100 km/day, the correlation coefficient between the speed values estimated from satellite/satellite SST MA





**Fig. 4. TIW westward propagation speed during the testing period.** The speed is estimated on the basis of the maximum detrended cross-correlation along the equator between the satellite MAs at the current time step and the satellite or forecasted MAs at the next time step (see Methods). **(A)** Speed at the whole 20° latitude band. The speed estimated from satellite/satellite SST MA pairs is denoted by deep brown dashed curve. The forecasted speed estimated from satellite/DL-forecasted SST MA pairs is denoted by green solid curve. The daily Niño3.4 index data are denoted by orange dotted curve. March 12, 2010 is the first day to be forecasted. **(B)** Distribution of speeds at 2° latitude bands estimated from satellite/satellite SST MA pairs and **(C)** distribution of forecasted speeds at 2° latitude bands estimated from satellite/DL-forecasted SST MA pairs. The white blanks represent outliers beyond the normal range of 0 to 100 km/day.

pairs and the Niño3.4 index values is  $-0.38$ . The  $P$  value, 95% confidence interval, and sample size are 0.0000,  $(-0.41, -0.35)$ , and 2882, respectively. The corresponding numbers for the forecasted speed values are  $-0.53$ , 0.0000,  $(-0.55, -0.50)$ , and 3235.

The TIW westward propagation speeds at 2° latitude bands are shown in Fig. 4 (B and C). We can see that the distributions estimated from the satellite/satellite and satellite/forecasted SST MA pairs are consistent with each other and have similar temporal fluctuations during TIW seasons, resembling the curves in Fig. 4A. Moreover, the speeds near the equator are higher than those at higher latitudes. All these calculation results are consistent with the previous findings because TIWs at different latitudes are dominated by different dynamical mechanisms and propagate at different speeds determined by equatorial wave processes (14, 16).

### Recursive forecasting

The DL model can also work recursively. In this way, we used the previous 12 satellite SSTs, the current satellite SST, and the forecasted SST (at the first recursive step) to generate the SST forecast at the next future time step (the second recursive step) and then used the previous 11 satellite SSTs, the current satellite SST, and the two

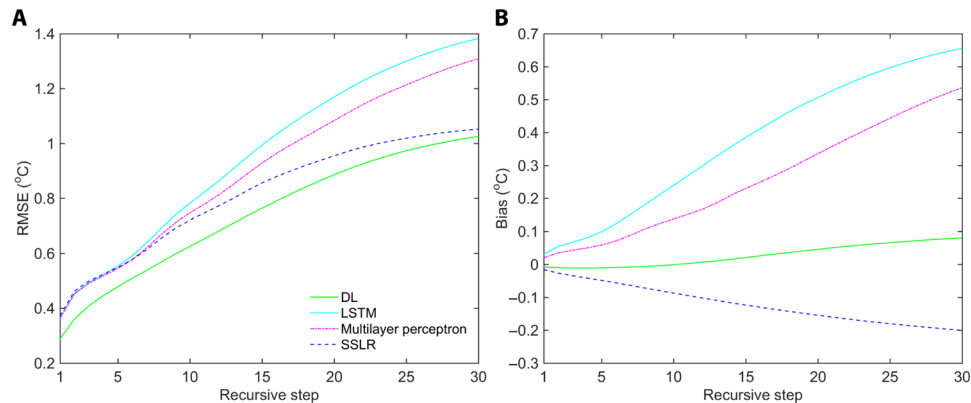
forecasted SSTs to generate the SST forecast at the further next future step (the third recursive step). Recursively, we made the SST forecasts at the subsequent time steps (the fourth, fifth, sixth, etc., recursive steps).

For the recursive forecasting, the global RMSE and bias of the DL model from recursive steps 1 to 30 (i.e., 5 to 150 days after the current time step) are given in Fig. 5. The accuracy of the DL model degrades with time because of a loss of the actual satellite SST input. It should be noted that after 14 recursive steps, there is no satellite SST at the model input. Even then, the RMSE is still smaller than  $0.80^{\circ}\text{C}$  at the 15th recursive step. The bias magnitude of the DL model is also lower than  $0.10^{\circ}\text{C}$  at the 30th recursive step.

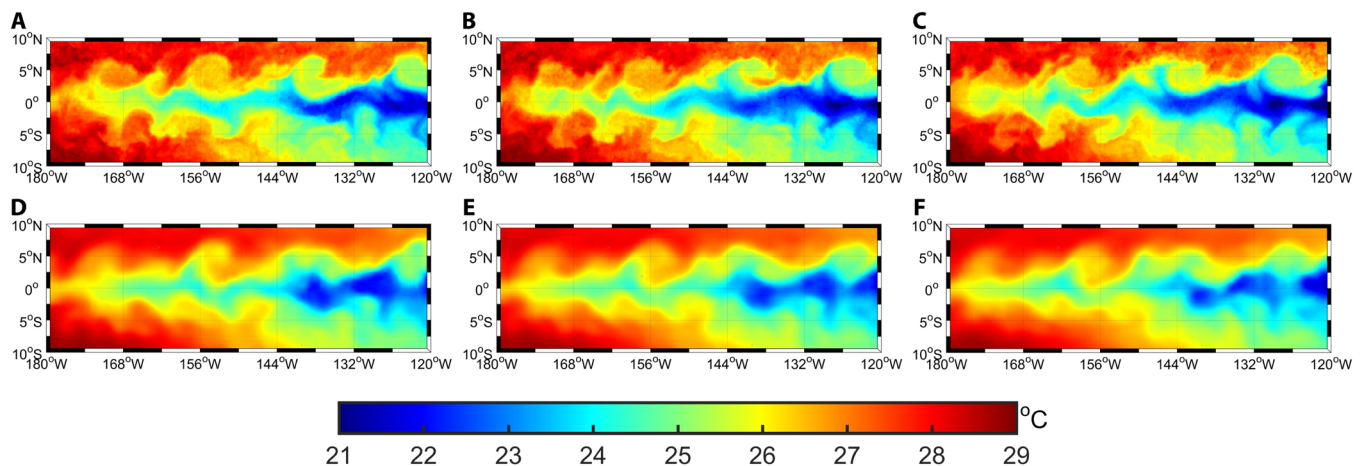
An instance of the recursively forecasted SST maps at the three subsequent time steps after the final time step in Fig. 3 is given in Fig. 6. We can see that although actual satellite SST maps at the model input become less due to more recursive steps, the DL model can still forecast the general westward motion of the TIWs.

### Comparison with other models

We comprehensively compared the DL model with the other three models, including analyzing their RMSE/bias temporal trends and



**Fig. 5. Trends of global RMSE and bias of DL model implemented recursively concerning the number of recursive steps. (A)** RMSE and **(B)** bias. In the recursive manner, the model-forecasted SST at a future time step is fed back to the model input to make the SST forecast at the next future time step. The recursive steps from 1 to 30 are 5 to 150 days after the current time step. After 14 recursive steps, there is no satellite SST map at the model input, and all input SST maps are forecasted SST maps. The global RMSE and bias of the DL model at each recursive step were calculated from the forecasting errors of all samples at all grids (green solid curve). The DL model was also compared with a long short-term memory (LSTM) model (cyan solid curve), a multilayer perceptron model (magenta dash-dotted curve), and a site-specific linear regression (SSLR) model (blue dashed curve). The detailed comparisons are given in the Supplementary Materials.



**Fig. 6. Satellite SST maps and those recursively forecasted by DL model at the three consecutive time steps (27 September and 2 and 7 October 2010) subsequent to the final time step in Fig. 3.** The final time step in Fig. 3 is the first recursive step here, and the three subsequent time steps are the second to fourth recursive steps. **(A to C)** Satellite SST maps at the three time steps and those forecasted in the recursive manner by the DL model **(D to F)**.

spatial distributions. Our DL model achieves higher accuracy than all three models. Moreover, unlike the DL model, the three models fail to forecast TIW motion (see section S1 for details).

## DISCUSSION

TIW is one of the essential and complicated oceanic phenomena, acquiring energy from nonlinear and chaotic hydrodynamic instabilities and propagating with large-shape distortions and deformations. The TIW-related SST field inextricably interacts with various oceanic, air-sea, biophysical, climate change, and processes. So, it is a critical factor when studying many relevant phenomena. Physical equation-based numerical modeling for TIWs requires not only high spatial resolution but also realistic parameterizations of the complex processes. As a result, substantial difficulties exist in realistically simulating TIWs. However, rules behind complicated oceanic

phenomena are deeply hidden in the time series of observation data itself and need to be dug up. Thus, the trend of quickly increasing volumes of satellite remote sensing big data and the powerful DL technology offer a promising alternative for modeling complicated oceanic phenomena with the difficulties circumvented, which is purely data-driven as we did for TIW in this study. However, this aspect has hardly been explored so far. Our DL model accurately forecasted the spatial-temporal variation of the SST field and the propagation of TIWs that are consistent with actual satellite observations. TIW speeds are higher on the equator than at higher latitudes.

The excellent performance of the DL model demonstrates the feasibility, effectiveness, and strong potential of the DNN technique for forecasting fields of oceanic phenomena. DL forecasts can be directly driven by actual observation data and avoid the complicated procedure of numerical forecast models, including various physical equations, model approximations/parameterizations, and huge

computational resource requirements. A DL model can generate accurate forecasts with prior information of very few physical parameters, e.g., only satellite SST data in our case. The model can be readily modified and implemented to process other gridded satellite remote sensing data such as sea surface wind, sea surface height, and chlorophyll- $\alpha$  field data. Most of the DL model is spent on the learning procedure of the DNN during the process of iteratively optimizing the adjustable weights. This procedure can easily speed up by combining with emerging hardware techniques such as Compute Unified Device Architecture. Once the DNN is trained and the bias correction map is constructed, the DL model can efficiently generate forecasts without any iteration and thus runs very fast. In our study, it took only approximately 1 min to complete the SST field forecast of the whole testing period of 9 years on a regular desktop computer. As a data-driven technique, sufficient data are always an essential requirement of the DNN technique whenever training a DNN or using it for ocean forecasting. This requirement and the excellent learning ability of DNNs match the ever-increasing volume of ocean satellite data generated in the remote sensing big data era.

While artificial neural network-based ocean forecasts have drawn a lot of attention (5, 7–9), some explorations were also made to atmospheric field forecasts very recently (35, 36). In the studies, the DL models were built and tested on simulation data of general circulation models and reanalysis data and achieved promising results (35, 36). With the improvement of network architecture, complicated phenomena can be more effectively forecasted by DL models, such as the newest advances in the forecast of El Niño–Southern Oscillation events (5) and the short-term predictions of tropical cyclone track and intensity (37, 38). A combination of DL technology and physical models is also essential (4), and recent advances have been well reviewed on the topic of artificial intelligence for remote sensing and numerical weather prediction applications, which are instructive and inspirational (38).

The growing error of the DL model with increasing recursive steps is to be reduced, if we want to recursively use the DL model for accurate forecasts at more future steps. Data assimilation could provide some cues in this aspect. In our study, we only used a single one DL model to make a forecast. Ensemble learning, which is widely used to boost performances of classification models in the field of machine learning, can generate an ensemble DL forecast model and should be one of the future endeavors in the DL ocean forecast field.

Our study extends the usage of ocean satellite data, enriches the application of DL technology in oceanography, and could promote the multidisciplinary research in this aspect, which is still in its infancy.

## METHODS

### Architecture and training of the DL model

As shown in Fig. 1, our DL model is composed of a multiscale DNN and a time-independent bias correction map. The DNN consists of four stacked composite layers, each of which has four cascaded convolutional layers. The SSTs in the area are generally in the range of 16° to 34°C, and we rescaled them to  $[-1, 1]$ . After the rescaling, the SST maps were down-sampled three times with a  $2 \times 2$  average pooling and then fed to the corresponding composite layers at different stack levels, which process the SST maps at different spatial resolutions. The lower the stack level, the higher the resolution. The output of each composite layer except the bottom composite layer

was up-sampled and fed to the higher-resolution composite layer at the lower stack level. The input of the DNN consists of 14 SST maps at the previous 13 and current time steps. The input SST map at the present time step is more correlative to the future SST map. Thus, in each composite layer, the input SST map at the current time step was also directly linked to the last convolutional layer along with the up-sampled output of the lower-resolution composite layer at the upper stack level. The activations of the first three convolutional layers of each composite layer are a rectified linear unit function due to its advantages such as cheaper computation cost and better gradient propagation (39). Except for the bottom composite layer, the activation of the last convolutional layer of each composite layer is a tanh function. The tanh function scales the output of each composite layer to  $[-1, 1]$  to match the input range of the higher-resolution composite layer where the output is fed after the up-sampling. The activation of the last convolutional layer of the bottom composite layer is a linear function. The linear function makes the DNN output unbounded. The four convolutional layers of each composite layer include 8, 16, 32, and 1 channels. The kernel sizes of the four convolutional layers of the top composite layer are all  $3 \times 3$ . For the other composite layers, the kernel sizes are  $5 \times 5$ ,  $3 \times 3$ ,  $3 \times 3$ , and  $5 \times 5$ .

For a general network layer dealing with map input and output, a site in the output map is connected with multiple sites, named the receptive field, in the input map. Thus, the value at the output site is only dependent on the values within the receptive field rather than the whole input map. In other words, the receptive field decides the extent of the spatial dependence of the value at the output site. For a convolutional layer, the values at the nearby input sites of an output site are weighted by the convolutional kernel parameters and then combined to obtain the value at the output site. Thus, the receptive field size is the same as the kernel size. For a composite layer consisting of cascaded convolutional layers, the receptive field can be easily found by recursion of the receptive fields of the convolutional layers (see fig. S2 for examples). The receptive field of the composite layer can be enlarged by down-sampling the input with an average pooling layer before feeding the input information to the composite layer. Then, the output can be up-sampled to recover the resolution. SST variations at different sites are highly related to each other due to the evolution of large-scale oceanic phenomena. Hence, when forecasting SST at a site, we used prior SST series at additional neighboring sites in the broader area centered at the forecast site. For this purpose, we enlarged the receptive field of the composite layer by adopting the multiscale scheme. After three times of down- and up-sampling among the four composite layers, the side length of the receptive field of the DNN increases approximately 12 times to 2900 km, which is sufficiently large to cover the entire study object, i.e., the TIW.

The samples of the SST map series during the training period are divided into two datasets with a ratio of 3:1 for training and validation. To avoid the receptive field of the DNN extending beyond the input area, we set the input area larger than that of the output (forecast) area. The following loss function is used for optimizing the DNN.

$$\text{Loss} = \sum_{k=1}^K \sum_{(m,n) \in \text{Grids}_{\text{output}}} \left( \text{SST}_{\text{output}}^{(k)}(m,n) - \text{SST}_{\text{true}}^{(k)}(m,n) \right)^2 \quad (1)$$

where  $\text{SST}_{\text{true}}^{(k)}(m,n)$  is the true SST (satellite SST) at the 15th (last) time step of the  $k$ th sample of the training or validation dataset at

the grid  $(m, n)$  of the output area, and  $SST_{\text{output}}^{(k)}(m, n)$  is the forecasted SST at the same time step output by the DNN.  $\text{Grids}_{\text{output}}$  denotes the set of the grids of the output area, and  $K$  is the total number of samples in the dataset. The parameters of the DNN were optimized over 2500 epochs by the Adam algorithm (40) to minimize the loss on the training dataset. The loss during the optimization of the DNN is shown in fig. S3. However, to avoid the overfitting of the optimization on the training dataset, we adopted the parameter values with the minimum loss on the validation dataset in the optimization process. The optimization was carried out on an NVidia Quadro M4000 GPU (8 GB of memory) with the Compute Unified Device Architecture technique. It took 129 epochs to achieve the optimal parameters with the minimum loss on the validation dataset, corresponding to approximately 114 min of computing time.

Convolutional layers are site-independent because their parameters, or their kernels, are shared across their input sites. Average pooling layers and up-sampling layers are also site-independent and have no optimizable settings. Thus, the DNN is site-independent. However, the local variation rule of the SST pattern in the forecasted area may have site differences due to differences in environment background, e.g., the approximately time-independent general SST trend of being higher/lower to the west/east in the equatorial Pacific Ocean. Therefore, a time-independent but site-dependent SST map is used to correct the DNN output in the DL model. This SST map is the spatial distribution of the bias correction of the DNN. The optimized DNN was applied to all the samples during the training period, and the bias was estimated at each grid to obtain the bias correction map (see the bottom right of Fig. 1). This map was added to the DNN output to get the forecasted SST map, as shown in Fig. 1.

After the DL model is developed, running it is very efficient; it only takes approximately 1 min to generate the SST pattern forecast of all the samples during the testing period on a regular desktop computer.

### Estimation of SST pattern zonal speed from the DL model

The westward propagation of the SST pattern was found in the MAs of the SST maps during TIW seasons. The signal overlies the approximately linear zonal trend of SST being colder in the eastern part of the forecast area and warmer in the western region. An instance of two zonal sequences of SST MAs at the longitudes of the grids of the forecast area and at two consecutive time steps is given in fig. S4. The first sequence (blue) is the MAs of the satellite SST map, and the second sequence (red) is those of the forecasted SST map after a one-time step (5 days). After the linear zonal trend of the SST MAs is removed, the westward propagating signal becomes more evident. The cross-correlations of the two sequences of detrended SST MAs are calculated at the discrete zonal lags. The discrete zonal lag with the maximum cross-correlation and its two neighboring discrete zonal lags are found. The cross-correlations at the three discrete zonal lags are interpolated by a quadratic curve, and the zonal lag with the quadratic curve reaching its peak is considered as the exact zonal lag with the maximum cross-correlation between the two detrended SST MA sequences. In mathematical form, this is

$$\text{lag}_{\text{exact}} = \frac{1}{2} \cdot \frac{y_1(\text{lag}_2^2 - \text{lag}_3^2) + y_2(\text{lag}_3^2 - \text{lag}_1^2) + y_3(\text{lag}_1^2 - \text{lag}_2^2)}{y_1(\text{lag}_2 - \text{lag}_3) + y_2(\text{lag}_3 - \text{lag}_1) + y_3(\text{lag}_1 - \text{lag}_2)} \quad (2)$$

where  $\text{lag}_1$ ,  $\text{lag}_2$ , and  $\text{lag}_3$  are the three discrete zonal lags, and  $y_1$ ,  $y_2$ , and  $y_3$  are the corresponding cross-correlations. Last, the propaga-

tion speed can be obtained by dividing the exact zonal lag by the time interval.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/29/eaba1482/DC1>

### REFERENCES AND NOTES

1. R. Legeckis, Long waves in the eastern equatorial Pacific Ocean: A view from a geostationary satellite. *Science* **197**, 1179–1181 (1977).
2. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
3. M. I. Jordan, T. M. Mitchell, Machine learning: Trends perspectives, and prospects. *Science* **349**, 255–260 (2015).
4. M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, Prabhat, Deep learning and process understanding for data-driven earth system science. *Nature* **566**, 195–204 (2019).
5. Y.-G. Ham, J.-H. Kim, J.-J. Luo, Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).
6. X. Li, B. Liu, G. Zheng, Y. Ren, S. Zhang, Y. Liu, L. Gao, Y. Liu, B. Zhang, F. Wang, Deep learning-based information mining from ocean remote sensing imagery. *Natl. Sci. Rev.*, nwa047 (2020).
7. K. Patil, M. C. Deo, Basin-scale prediction of sea surface temperature with artificial neural networks. *J. Atmos. Ocean. Technol.* **35**, 1441–1455 (2018).
8. Q. Zhang, H. Wang, J. Dong, G. Zhong, X. Sun, Prediction of sea surface temperature using long short-term memory. *IEEE Geosci. Remote Sens. Lett.* **14**, 1745–1749 (2017).
9. S. G. Aparna, S. D'Souza, N. B. Arjun, Prediction of daily sea surface temperature using artificial neural networks. *Int. J. Remote Sens.* **39**, 4214–4231 (2018).
10. L. Qiao, R. H. Weisberg, Tropical instability wave kinematics: Observations from the Tropical Instability Wave Experiment. *J. Geophys. Res.* **100**, 8677–8693 (1995).
11. D. B. Chelton, F. J. Wentz, C. L. Gentemann, R. A. de Szoeke, M. G. Schlax, Satellite microwave SST observations of transequatorial tropical instability waves. *Geophys. Res. Lett.* **27**, 1239–1242 (2000).
12. R. F. Contreras, Long-term observations of tropical instability waves. *J. Phys. Oceanogr.* **32**, 2715–2722 (2002).
13. C. S. Willett, R. R. Leben, M. F. Lavin, Eddies and tropical instability waves in the eastern tropical Pacific: A reviews. *Prog. Oceanogr.* **69**, 218–238 (2006).
14. J. M. Lyman, G. C. Johnson, S. Kessler, Distinct 17- and 33-day tropical instability waves in subsurface observations. *J. Phys. Oceanogr.* **37**, 855–872 (2007).
15. M. Jochum, M. F. Cronin, W. S. Kessler, D. Shea, Observed horizontal temperature advection by tropical instability waves. *Geophys. Res. Lett.* **34**, L09604 (2007).
16. T. Lee, G. Lagerloef, M. M. Gierach, H.-Y. Kao, S. Yueh, K. Dohan, Aquarius reveals salinity structure of tropical instability waves. *Geophys. Res. Lett.* **39**, L12610 (2012).
17. J. N. Moum, R.-C. Lien, A. Perlin, J. D. Nash, M. C. Gregg, P. J. Wiles, Sea surface cooling at the equator by subsurface mixing in tropical instability waves. *Nat. Geosci.* **2**, 761–765 (2009).
18. C. Liu, A. Köhl, Z. Liu, F. Wang, D. Stammer, Deep-reaching thermocline mixing in the equatorial Pacific cold tongue. *Nat. Commun.* **7**, 11576 (2015).
19. R. Inoue, R.-C. Lien, J. N. Moum, R. C. Perez, M. C. Gregg, Variations of equatorial shear, stratification and turbulence within a tropical instability wave cycle. *J. Geophys. Res. Oceans* **124**, 1858–1875 (2019).
20. R.-H. Zhang, A. J. Busalacchi, Rectified effects of tropical instability wave (TIW)-induced atmospheric wind feedback in the tropical Pacific. *Geophys. Res. Lett.* **35**, L05608 (2008).
21. J. A. Yoder, S. G. Ackleson, R. T. Barber, P. Flament, W. M. Balch, A line in the sea. *Science* **371**, 689–692 (1994).
22. W. Evens, P. G. Strutton, F. P. Chavez, Impact of tropical instability waves on nutrient and chlorophyll distributions in the equatorial Pacific. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **56**, 178–188 (2009).
23. F. Tian, R.-H. Zhang, X. Wang, A positive feedback onto ENSO due to tropical instability wave (TIW)-induced chlorophyll effects in the Pacific. *Geophys. Res. Lett.* **46**, 889–897 (2019).
24. S.-L. An, Interannual variations of the tropical ocean instability wave and ENSO. *J. Climate* **21**, 3680–3686 (2008).
25. Y. Imada, M. Kimoto, Parameterization of tropical instability waves and examination of their impact on ENSO characteristics. *J. Climate* **25**, 4568–4581 (2012).
26. R. M. Holmes, S. McGregor, A. Santoso, M. H. England, Contribution of tropical instability waves to ENSO irregularity. *Climate Dynam.* **52**, 1837–1855 (2019).
27. W. Hsieh, B. Tang, Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Am. Meteor. Soc.* **79**, 1855–1870 (1998).



28. S. J. Murphy, R. Washington, T. E. Downing, R. V. Martin, G. Ziervogel, A. Preston, M. Todd, R. Butterfield, J. Briden, Seasonal forecasting for climate hazards: Prospects and responses. *Nat. Hazards* **23**, 171–196 (2001).
  29. S. N. Londhe, V. Panchang, One-day wave forecasts based on artificial neural networks. *J. Atmos. Ocean. Technol.* **23**, 1593–1603 (2006).
  30. X. Zeng, Y. Li, R. He, Predictability of the loop current variation and eddy shedding process in the Gulf of Mexico using an artificial neural network approach. *J. Atmos. Ocean. Technol.* **32**, 1098–1111 (2015).
  31. J. Chi, H.-C. Kim, Prediction of Arctic sea ice concentration using a fully data driven deep neural network. *Remote Sens.* **9**, 1305 (2017).
  32. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention 2015*, Munich, Germany, 5 to 9 October 2015.
  33. L. Qiao, R. H. Weisberg, Tropical instability wave energetics: Observations from the Tropical Instability Wave Experiment. *J. Phys. Oceanogr.* **28**, 345–360 (1998).
  34. S. Masina, S. G. H. Philander, A. B. G. Bush, An analysis of tropical instability waves in a numerical model of the Pacific Ocean: 2. Generation and energetics of the waves. *J. Geophys. Res.* **104**, 29637–29661 (1999).
  35. S. Scher, G. Messori, Weather and climate forecasting with neural networks: Using general circulation models (GCMs) with different complexity as a study ground. *Geosci. Model Dev.* **12**, 2797–2809 (2019).
  36. J. A. Weyn, D. R. Durran, R. Caruana, Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J. Adv. Model. Earth Syst.* **11**, 2680–2693 (2019).
  37. K. A. Cloud, B. J. Reich, C. M. Rozoff, S. Alessandrini, W. E. Lewis, L. D. Monache, A feed forward neural network based on model output statistics for short-term hurricane intensity prediction. *Wea. Forecasting* **34**, 985–997 (2019).
  38. S.-A. Boukabara, V. Krasnopolsky, J. Q. Stewart, E. S. Maddy, N. Shahroudi, R. N. Hoffman, Leveraging modern artificial intelligence for remote sensing and NWP. *Bull. Am. Meteorol. Soc.* **100**, ES473–ES491 (2019).
  39. X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in *14th International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, 11 to 13 April 2011.
  40. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in *International Conference on Learning Representations 2015*, San Diego, CA, 7 to 9 May 2015.
- Acknowledgments:** We would like to thank Remote Sensing Systems for providing the satellite remote sensing SST products and KNMI Climate Explorer for providing the Niño3.4 data. **Funding:** This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant nos. XDB42000000 and XDA19090103); the Key R&D Project of Shandong Province (grant no. 2019JZZY010102); the National Natural Science Foundation of China (grant nos. 41676167 and 41776183); the Key Deployment Project of Center for Ocean Mega-Science, CAS (grant no. COMS2019R02); the CAS Program (grant no. Y9KY04101L); and the Project of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (grant no. SOEDZZ2003). **Author contributions:** G.Z. and X.L. devised and conducted the original study. G.Z., X.L., and R.-H.Z. discussed and analyzed the model forecasts. All the authors contributed to writing the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The Remote Sensing Systems blended daily global SST products are available at [www.remss.com/measurements/sea-surface-temperature](http://www.remss.com/measurements/sea-surface-temperature); the KNMI Climate Explorer daily Niño3.4 index data are available at [https://climexp.knmi.nl/getindices.cgi?WMO=NCEPData/nino34\\_daily&STATION=NINO3.4&TYPE=i&id=someone@somewhere&NPYEAR=366](https://climexp.knmi.nl/getindices.cgi?WMO=NCEPData/nino34_daily&STATION=NINO3.4&TYPE=i&id=someone@somewhere&NPYEAR=366). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.
- Submitted 8 November 2019  
Accepted 1 June 2020  
Published 15 July 2020  
10.1126/sciadv.aba1482
- Citation:** G. Zheng, X. Li, R.-H. Zhang, B. Liu, Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* **6**, eaba1482 (2020).