AUTHOR QUERY FORM

AIP Publishing	**Journal: Chaos** **Article Number: CHA22-AR-DDCS2022-00825**	**Please provide your responses and any corrections by annotating this PDF and uploading it to AIP's eProof website as detailed in the Welcome email.**

Dear Author,

Below are the queries associated with your article; please answer all of these queries before sending the proof back to AIP.

Article checklist: In order to ensure greater accuracy, please check the following and make all necessary corrections before returning your proof.
1. Is the title of your article accurate and spelled correctly?
2. Please check affiliations including spelling, completeness, and correct linking to authors.
3. Did you remember to include acknowledgment of funding, if required, and is it accurate?

Location in article	Query / Remark: click on the Q link to navigate to the appropriate spot in the proof. There, insert your comments as a PDF annotation.
Q1	Please check that the author names are in the proper order and spelled correctly. Also, please ensure that each author's given and surnames have been correctly identified (given names are highlighted in red and surnames appear in blue).
Q2	Please define SMOE at first occurrence.
Q3	Please define CCN at first occurrence.
Q4	Please include the dataset reference in the reference list and update the reference in the sentence beginning "The data that support…."
Q5	We were unable to locate a digital object identifier (doi) for Refs. Arrieta *et al.* (2020), Butterworth *et al.* (1930), Duan *et al.* (2009), L'Heureux *et al.* (2017), Srivastava *et al.* (2014), Yan *et al.* (2020), and Zhao and Di Lorenzo (2020). Please verify and correct author names and journal details (journal title, volume number, page number, and year) as needed and provide the doi. If a doi is not available, no other information is needed from you. For additional information on doi's, please select this link: http//www.doi.org/.
Q6	Please provide publisher name for Refs. Montavon *et al.* (2019) and Selvaraju *et al.* (2017).
Q7	The resolution of Figs 2,4,7,8,10,13 and 15 is low. If you are not satisfied with the way the figures appear in the proof, please provide new figure files of higher resolution.
	Please confirm ORCIDs are accurate. If you wish to add an ORCID for any author that does not have one, you may do so now. For more information on ORCID, see https://orcid.org/. G. Lancia–0000-0001-6185-7081 I. J. Goede C. Spitoni H. Dijkstra

	Please check and confirm the Funder(s) and Grant Reference Number(s) listed: Netherlands Science Foundation, OCENW.M20.277 Please add any additional funding sources not stated above:

Thank you for your assistance.

Physics captured by data-based methods in El Niño prediction

Cite as: Chaos **32**, 000000 (2022); doi: 10.1063/5.0101668
Submitted: 2 June 2022 · Accepted: 19 September 2022 ·
Published Online: ■■ ■■ 2022

G. Lancia,[1,a)] I. J. Goede,[2] C. Spitoni,[1,3] and H. Dijkstra[2,3]

AFFILIATIONS

[1]Department of Mathematics, Utrecht University, Budapestlaan 6, 3584 CD Utrecht, Netherlands
[2]Institute for Marine and Atmospheric Research Utrecht, Department of Physics, Utrecht University, Princetonplein 5, 3584 CC Utrecht, Netherlands
[3]Center for Complex Systems Studies, Department of Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, Netherlands

Note: This article is part of the Focus Issue, Theory-informed and Data-driven Approaches to Advance Climate Sciences.
[a)]**Author to whom correspondence should be addressed:** g.lancia@uu.nl

ABSTRACT

On average once every four years, the Tropical Pacific warms considerably during events called El Niño, leading to weather disruptions over many regions on Earth. Recent machine-learning approaches to El Niño prediction, in particular, Convolutional Neural Networks (CNNs), have shown a surprisingly high skill at relatively long lead times. In an attempt to understand this high skill, we here use data from distorted physics simulations with the intermediate-complexity Zebiak–Cane model to determine what aspects of El Niño physics are represented in a specific CNN-based classification method. We find that the CNN can adequately correct for distortions in the ocean adjustment processes, but that the machine-learning method has far more trouble ~~to~~ dealing with distortions in upwelling feedback strength.

Tropical Pacific can periodically be subjected to an irregular variation in sea surface temperature (SST), affecting the climate over many regions on Earth. In the last decade, deep learning techniques, in specific Convolutional Neural Networks (CNNs), have shown to be peculiarly accurate in El Niño predictions, even at long lead times. In order to give a deeper understanding and an interpretation of this high skill of CNN, we make use of data from distorted physics simulations to determine what aspects of El Niño physics can be captured and recognized in a specific CNN-based classification method. We find that the CNN can capture the wave adjustment and feedback process, but that the deep learning method has far more trouble ~~to~~ dealing with distortions in upwelling feedback strength.

I. INTRODUCTION

Interannual climate variability is strongly dominated by the El Niño-Southern Oscillation (ENSO) in the Tropical Pacific. During an El Niño, the positive phase of ENSO, sea surface temperatures in the eastern Pacific increase by a few degrees with respect to seasonally averaged values; the oscillation phase opposite to El Niño is La Niña, with a colder eastern Pacific. A much used measure of the state of ENSO is the NINO3.4 index, which is the area-averaged Sea Surface Temperature (SST) anomaly [i.e., deviation with respect to the mean seasonal cycle (SC)] over the region 170°W–120°W × 5°S–5°N. El Niño events typically peak in December, occur every two to seven years, and their strength varies irregularly on decadal time scales. The spatial pattern of ENSO variability is often determined from principal component analysis (Preisendorfer, 1988), detecting patterns of maximal variance. At least two different types of El Niño events exist (Kug *et al.*, 2009; and Zhang *et al.*, 2019), with the largest temperature anomalies either in the eastern Pacific (Eastern Pacific or EP El Niño's) or near the dateline (Central Pacific or CP El Niño's).

As ENSO has distinct influences on the climate around the globe through well-known teleconnections (Diaz *et al.*, 2001), skillful predictions of up to a one year lead time are desired to be able to mitigate the effects (Balmaseda *et al.*, 1995). For ENSO predictions, often the Oceanic Niño Index (ONI) is used, which is defined as the three-month running mean of the NINO3.4 index. Both statistical models (those capturing behavior of past events) and dynamical models (i.e., those based on the underlying physical conservation laws) are used for El Niño prediction (Latif, 1998; Chen and Cane, 2008; Barnston *et al.*, 2012; Saha *et al.*, 2014; Timmermann *et al.*, 2018; Tang *et al.*, 2018; and Barnston *et al.*, 2019). El Niño events are difficult to predict as they have an irregular occurrence and each time have a slightly different development (McPhaden *et al.*, 2015; and Timmermann *et al.*, 2018). Many ENSO prediction evaluation studies (Barnston *et al.*, 2012; and L'Heureux *et al.*, 2017) have shown that dynamical models do better than statistical models, although there are exceptions (Newman and Sardeshmukh, 2017). When initialized before the boreal spring, most models perform much worse than when initialized in summer. The latter notion has been indicated by the spring predictability barrier problem (McPhaden, 2003).

ENSO theory (Neelin *et al.*, 1998) provides a framework to understand the existence of such predictability barriers. The ENSO phenomenon is thought to be an internal mode of the coupled equatorial ocean-atmosphere system which can be self-sustained or excited by small-scale processes, often considered as noise (Fedorov *et al.*, 2003). Bjerknes' feedbacks are central in the amplification of SST anomalies, whereas equatorial ocean wave processes provide a delayed negative feedback and are responsible for the time scale of ENSO. The interactions of the internal mode and the external seasonal forcing can lead to chaotic behavior through nonlinear resonances (Tziperman *et al.*, 1994; and Jin *et al.*, 1994). On the other hand, the dynamical behavior can be strongly influenced by noise, in particular, westerly-wind bursts (Lian *et al.*, 2014). During boreal spring and summer, the Pacific climate system is most susceptible to perturbations leading to predictability barriers (Latif and Barnett, 1994). The growth of perturbations from a certain initial state has been investigated in detail from a much used intermediate-complexity model, the Zebiak–Cane (ZC) model (Zebiak and Cane, 1987). Applying the methodology of optimal modes (Mu *et al.*, 2007; Duan *et al.*, 2009; and Yu *et al.*, 2012), it was indeed shown that spring is the most sensitive season for EP El Niños and likewise summer for CP El Niños (Tian and Duan, 2015; and Hou *et al.*, 2019).

Deep learning methods (DLMs) are powerful statistical models, which have now been used in a wide range of applications such as speech recognition and image reconstruction (Goodfellow *et al.*, 2016). These methods include feed-forward Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), Reservoir Computers (RCs), and ~~Convolutional Neural Networks (CNNs);~~ over quite some time now, DLMs have been applied to El Niño prediction (Dijkstra *et al.*, 2019). The current work is motivated by the high El Niño prediction skill of two types of DLMs. First, in Ham *et al.* (2019), CNNs were trained on model data from the Climate Model Intercomparison Project, phase 5 (CMIP5) using transfer learning and subsequently trained on reanalysis data. The CNN-based scheme shows a better forecasting skill than most dynamical models and this forecast skill remains high up to lead times of about 17 months. It is also able to successfully predict the type of El Niño (CP or EP) patterns that develop. Second, in Petersik and Dijkstra (2020), deep ensemble methods (Lakshminarayanan, 2017), in particular, Gaussian Density Neural Networks (GDNNs) and Quantile Regression Neural Networks (QRNNs), were used in ENSO prediction. These methods also give a skillful model for the long-lead time prediction of the ONI (and its uncertainty) using a relatively small predictor set.

At the moment, there is an enormous effort to understand the performance of DLMs generally referred to as explainable AI (Arrieta *et al.*, 2020). The research described above shows that DLMs are a very promising tool in ENSO prediction that can provide useful skills of El Niño forecasts beyond the predictability barriers. The intriguing question is now what the DLMs capture of the ENSO physics contained in the data. Addressing this question is precisely the focus of this paper. We will approach this issue using the Zebiak–Cane model, which is also routinely used for ENSO prediction. The novel aspect of this work is that we use so-called distorted physics experiments where different physical processes (such as equatorial wave dynamics and ocean-atmosphere feedbacks) are perturbed. Using saliency analyses, determining which input variables contribute most to the prediction skill, we then aim to determine what part of the ENSO dynamics is represented by the DLM.

II. MODELS AND METHODS

A. ENSO model

The Zebiak–Cane (ZC) model (Zebiak and Cane, 1987) represents the coupled ocean-atmosphere system on an equatorial β-plane in the equatorial Pacific. In this model, a shallow-water ocean component is coupled to a steady shallow-water (Gill, 1980) atmosphere component (Fig. 1). The atmosphere is driven by heat fluxes from the ocean, depending linearly on the anomaly of the sea surface temperature T with respect to a radiative equilibrium temperature T_0. We use the numerically implicit fully-coupled version of this model, developed in van der Vaart *et al.* (2000) and slightly extended in Feng and Dijkstra (2017). In this version, the zonal wind stress τ^x is written as

$$\begin{aligned} \tau^x &= \tau^x_{ext} + \tau^x_c, \\ \tau^x_{ext} &= -\tau_0\, e^{-\frac{1}{2}\left(\frac{y}{L_a}\right)^2}. \end{aligned} \tag{1}$$

Here, the external part τ^x_{ext} represents a weak ($\tau_0 \sim 0.01$ Pa) easterly wind stress due to the Hadley circulation, L_a is the atmospheric Rossby deformation radius and y is the meridional coordinate. The zonal wind stress τ^x_c is proportional to the zonal wind from the atmospheric model which, in turn, depends on sea surface temperature.

As shown in van der Vaart *et al.* (2000), the parameter measuring the strength of all ocean-atmosphere coupled feedbacks is the coupling strength μ. When $\mu < \mu_c$, where μ_c indicates a critical value, the Tropical Pacific climatology (a stationary state of the model) is stable. However, if the coupling strength exceeds the critical value μ_c, a supercritical Hopf bifurcation occurs and sustained oscillations occur with a period of approximately four years. A seasonal cycle is included in the model by varying μ over time with a specific amplitude $\Delta\mu$ and with an annual period.

Apart from the coupled ocean-atmosphere processes, ENSO is also affected by fast processes in the atmosphere, such as westerly-wind bursts. These processes are considered as noise in the ZC model. The representation of atmospheric noise in the model is similar to that in Feng and Dijkstra (2017), where the westerly-wind

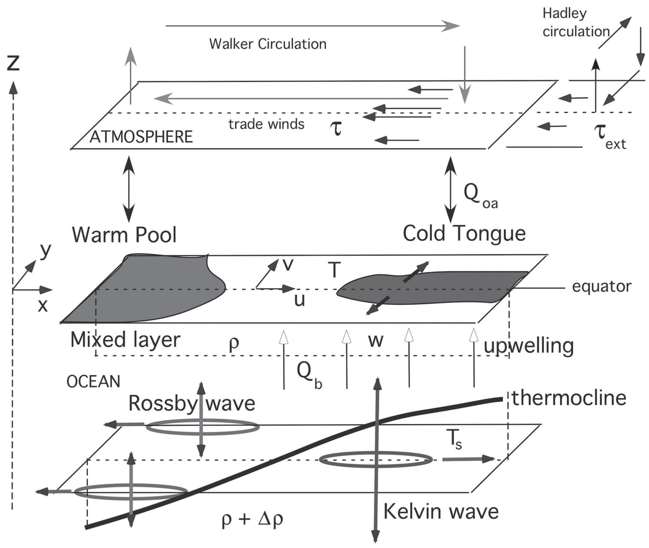

FIG. 1. Schematic of the Zebiak–Cane model, where a shallow-water ocean model is coupled to a shallow-water atmosphere model through a mixed-layer ocean model with temperature T. The ocean-atmosphere coupling involves a heat flux Q_{oa} and a wind-stress vector τ.

bursts are represented by one Empirical Orthogonal Function pattern (with the associated principle component fitted to an AR(1) process) in the zonal wind stress. The observation-based dataset in Feng and Dijkstra (2017) contains weekly patterns of this wind-stress noise. In the ZC model, we randomly add one of such patterns at each time step (of a week) to the zonal wind stress. The effect of the noise on the model behavior depends on whether the model is in the super- or sub-critical regime (i.e., whether μ above or below μ_c). If $\mu < \mu_c$, the noise excites the ENSO mode, causing irregular oscillations. In the supercritical regime, the cycle of approximately four years is still present, but the noise causes an irregular amplitude of ENSO variability.

While the ZC model is used for ENSO predictions, it also has its limitations as it cannot capture either tropical basin interactions (e.g., Atlantic and Indian Oceans) or tropical-extratropical interactions (it described only the dynamics of the Pacific). The model also cannot represent adequately a Tropical Pacific seasonal cycle, and, hence, such a seasonal cycle is prescribed in the model.

B. Distorted physics simulations

The advantage of the ZC model is that the behavior of the model can be connected to the physical processes in a very transparent way (Jin, 1997). In the distorted physics approach, we define a "truth" by a reference simulation, using an external seasonal cycle, prescribed noise in the wind-stress, and parameter settings such as in Feng and Dijkstra (2017). Next, in subsequent distorted-physics simulations, we change the representation of physical processes in the model by varying parameters. We will focus on the main processes setting the time scale and amplitude of ENSO.

An important memory component in the Tropical Pacific climate system is the ocean adjustment to changes in the atmospheric forcing. This is accomplished by equatorial wave dynamics and best described by a basin mode response, where the basin mode consists of a sum of one Kelvin and multiple Rossby waves. In the SST-mixed ocean dynamics mode framework behind ENSO variability (Neelin *et al.*, 1998), the adjustment is crucial for the timing of El Niño events. It plays also a crucial role in the recharge/discharge oscillator view of ENSO (Jin, 1997), where the equatorial heat content is varied, usually measured by the warm water volume (WWV) in observations. The temporal aspects of the adjustment can be controlled in the Zebiak–Cane model by putting a coefficient δ before the time derivatives of the ocean momentum equations (Neelin, 1991). In the extreme case where the time derivative is effectively zero ($\delta = 0$), the so-called "fast-wave" limit is reached.

Three of the most important positive Bjerknes' feedbacks are the thermocline feedback, the zonal advection feedback, and the upwelling (or Ekman) feedback (Dijkstra, 2005). The relative magnitude of these feedbacks determines which spatial SST perturbation patterns are amplified. In addition, the feedbacks determine also the mean state and seasonal cycle of the tropical Pacific climate state (Dijkstra and Neelin, 1995). Specific feedback strengths can be changed in the ZC model by varying the mean thermocline depth (thermocline feedback), the mean zonal temperature gradient (zonal advection feedback), or Ekman friction (upwelling feedback). We will concentrate on the latter feedback, affecting the amplitude of ENSO and which can be changed in the ZC model by adjusting the parameter δ_s.

Hence, we have already a good idea of how the behavior of the model is distorted by varying the parameters δ and δ_s. Now, δ is an artificial parameter enabling the variation of the equatorial wave speeds and in the results below, we vary it from 0.5 to 1.5. The upwelling feedback strength δ_s is quite an uncertain parameter in the ZC model and we vary it over the range $\delta_s = 0.1$ to $\delta_s = 0.6$, which is a plausble range, where an adequate mean state and variability are obtained (van der Vaart *et al.*, 2000). In the approach below, we are interested in whether a CNN is able to capture ENSO dynamics adequately when trained with data from distorted model simulations.

C. CNN approach

Due to their versatility and peculiarity in solving binary and multi-labels classification tasks by capturing and recognizing the discerning patterns of the input data, CNNs (Convolutional Neural Networks) can represent a powerful method for making forecasting of ENSO events with lead times of up to one and a half years (Ham *et al.*, 2019) or for solving a binary classification problem in hybrid models with high complexity multi-resolution input data (Yan *et al.*, 2020). Unlike more sophisticated and popular ANNs like CNN-LSTM and ConvLSTM, the predictions provided by the CNN can be made explainable by means of saliency maps (Zhou *et al.*, 2016; Selvaraju *et al.*, 2017; Adebayo *et al.*, 2018; Montavon *et al.*, 2019; and Mundhenk *et al.*, 2019) that allow us to outline the spatial locations of those signal patterns that mainly contribute to making the CNN give the classes of output. Therefore, CNNs represent the perfect choice for classifying the occurrence of ENSO

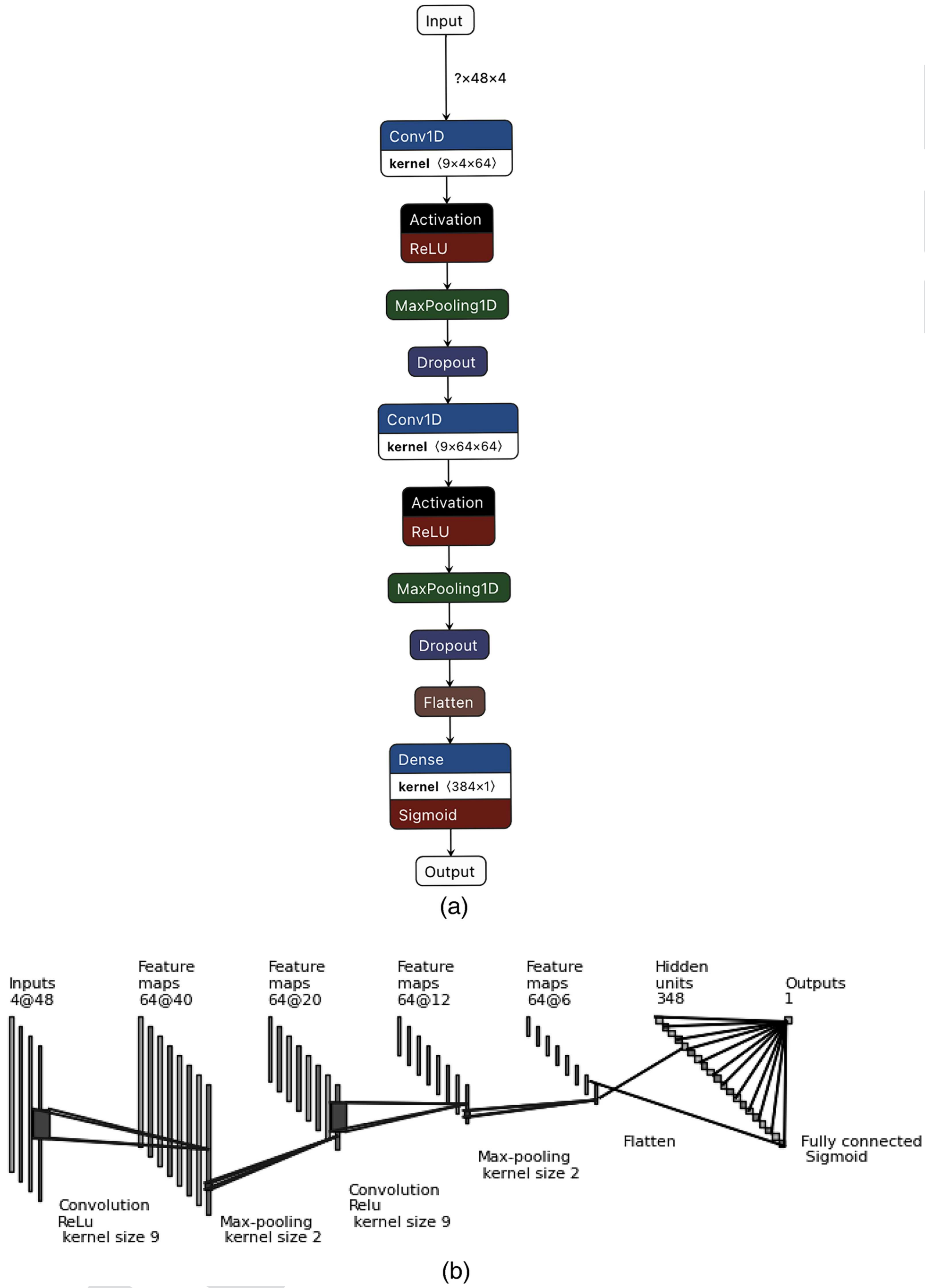

FIG. 2. Schematic illustrations of the CNN model; flow diagram (a) and the hidden layers (b).

events in ZC simulations and investigating in detail on which features contained in data can lead to highly accurate predictions. To leverage the basic feature of the CNN of encoding the sequentiality of the patterns contained in the input data, we feed the CNN with simulated time series obtained via the Zebiak–Cane model. This synthetic dataset describes the temporal evolution in the NINO3.4 region of some physical observables of interest as the thermocline depth, the sea surface temperature, the wind speed, and the wind-stress noise. The extraction of the instances from the ZC simulations, therefore, consists in slicing the synthetic time series along the time domain, i.e., the set of time series is chunked in a sequence of overlapping time windows of 48 months and stride 1. As a result, each single instance is a tensor of rank 2 whose dimensions are the time-length (48 pixels, sampling frequency one month) and the number of time-series features (the four physical observables of interest). The labeling of the instances is performed by equipping each instance with the corresponding ONI-index value and so we label one ENSO event whenever the ONI-index value is greater than 0.5 (El Niño event) or lower than −0.5 (La Niña event). The input instances are then pre-processed via standardization (each feature has now zero-mean and unit variance) and divided into training set and test set; we validate the CNN model by means of the fivefold cross validation. Therefore, we evaluate the AUC (Area Under the Curve) of the receiver operating characteristic curve on each fold; the mean value and the standard error mean will provide the degree of accuracy of the CNN model and its error, respectively. The design of our CNN is quite standard and it is composed by the sequence of one convolutional layer (64 kernels, size 9) with a Rectified Linear Unit (ReLu) activation function, followed by a maxpooling Layer with pooling size 2 (Fig. 2). Dropout layers (Srivastava *et al.*, 2014) with a dropout rate of 0.50 are also employed to reduce overfitting, but no stride is applied during the convolutions. After repeating two times this block of hidden layers, the resulting feature map is flattened via a flattened layer; the final fully-connected layer with the sigmoid activation function returns the output of the CNN. During the training phase, the ADAM (Kingma and Ba, 2014) algorithm is used as an optimizer for the binary-cross entropy loss function; the batch size and learning rate are set equal to 128 and 0.005, respectively. The ~~SMOE~~ scale method (Mundhenk *et al.*, 2019) is a robust statistical measure of the activation values of CNNs arising at different spatial locations (temporal locations in the domain of our instances). This statistics can be used to construct robust saliency maps that appear to be much more efficient and computationally faster than popular gradient methods. More specifically, this method estimates the saliency of the input data at each temporal location; the saliency values are returned as a score laying in the range [0, 1]. Therefore, the closest is one score to the unit value, the more is the saliency attributed to its temporal location. We, therefore, exploit the capability of SMOE scale method to detect those patterns and their spatial domains that mostly indicate the approaching or the occurring of the ENSO events. Thus, we proceed with the analysis of the profile of the saliency maps in order to evince possible analogies and differences between the patterns learnt during the training phase and the patterns contained in the test dataset. In order to complete the analysis provided by the SMOE scale method, we even look at how the predictions can change when only a spectral subband of the input instances is propagated through the hidden layers. With this approach, we aim to investigate how oscillations occurring under a specific regime can really be a basic aspect of the prediction provided by the CNN. Therefore, we can progressively apply a digital Butterworth filter (Butterworth *et al.*, 1930; and Hamming, 1998) of order 3 as either a bandpass filter or low-pass filter to smooth the input instance. The ensemble of bandpass filters is designed to cover the whole spectral domain of any input instance and be non-overlapping at the same time and, thus, we impose the cutoff frequencies of each filter to be in ratio 1:2. This means that, starting from the Nyquist frequency ν_0, the first digital filter will have its frequency band in $\left[\frac{\nu_0}{2}, \nu_0\right]$, the second one in $\left[\frac{\nu_0}{4}, \frac{\nu_0}{2}\right]$, and so on. Again, when considering the low-pass digital filters, we will choose the cut-off frequency according to a dyadic scale, i.e., the first filter will have cut-off frequency ν_0, the second one $\frac{\nu_0}{2}$, the third one $\frac{\nu_0}{4}$, and so on. The full list of bandwidths (in periods) and cutoff frequencies is reported in Table I. Note that we will apply these digital filtering techniques by repeating the same fivefold cross validation with metrics AUC, as we do in the model validation; the CNN architecture will not be altered during this step. Hence, by means of this approach, we aim to reveal which time scale is dominant in those patterns that characterize the ENSO events (e.g., a slow oscillating trends against rapid oscillating deviations), i.e., we make an effort to understand how the periodicity of the time-series features is an essential characteristic of data that the CNN captures for solving the classification task and how a distortion of it can give rise to a decrease in the CNN capability of classifying the events El Niño and La Niña.

TABLE I. Frequency bands and cut-off frequencies for the bandpass and low-pass digital filters, respectively.

Frequency bands (period in months)	Cut-off period (months)
[2, 4)	2
[4, 8)	4
[8, 16)	8
[16, 32)	16
[32, 48)	32

III. RESULTS

A. Distorted physics

The model experiments broadly consist of two steps: first, the ZC model is run for standard parameter values to produce reference case data; and then it is run again but for a range of values around the standard parameter value (shown in Table II) to get the distorted data. This ultimately results in three different kinds of datasets: reference case, distorted wave speed, and distorted upwelling feedback. There are no simulations where more than one parameter is distorted at the same time. In the second step, the distorted datasets are used as training data for the DLMs whose performance is then determined by using the reference case as the test set. As a consistency check, the DLMs are also trained on the reference case data and then tested on reference case data. This should produce the highest performance because the DLMs are tested on data they have already seen.

B. Equatorial wave dynamics: Saliency maps

Time series of the ONI for the different δ values, as computed from the ZC model are shown in Fig. 3. Changing the δ value causes the amplitude of the oscillation to become much smaller for $\delta < 1$, so much even that by definition only ENSO neutral conditions ($-0.5 < ONI < 0.5$) are present. Increasing δ above the reference

TABLE II. Parameter settings of the ZC model used to generate the data used in the distorted physics experiments with the parameter step size shown within brackets. Parameter ranges are chosen to cover roughly a 50% increase and decrease compared to the reference value, step size is chosen to get around 10 points within this range. The parameters are from left to right: coupling strength μ, wave speed parameter δ, and upwelling feedback parameter δ_s. The value of $\mu = 2.7$ is subcritical in the ZC model.

Effect	μ	δ	δ_s
Distorted wave speed	2.7	0.5–1.5 (0.1)	0.3
Distorted wave speed	2.7	0.5–1.5 (0.1)	0.3
Distorted upwelling feedback	2.7	1.0	0.1–0.6 (0.05)
Distorted upwelling feedback	2.7	1.0	0.1–0.6 (0.05)
Reference	2.7	1.0	0.3

value of 1.0 initially leads to an increase in the oscillation amplitude and it then decreases again for higher values of δ. This is expected because the ENSO period depends on the speed of Rossby and Kelvin waves crossing the Pacific basin. In the study of the classification performance of the CNN, we take a prediction lead time of 9 months. The propagation of the δ-distorted data through the CNN can lead to substantial changes when testing the accuracy of the model on the reference data. By construction, the AUC score [Fig. 4(a)] attains excellent results at $\delta = 1.0$ (AUC 0.94) as the CNN is trained on the reference data. The AUC scores tend to remain relatively high (peak of AUC 0.91 at $\delta = 0.8$) as the δ parameter is slightly decreased from its reference value. Instead, as δ is reduced up to value 0.5, we can observe a severe degradation of the accuracy with respect to the reference case; from $\delta = 0.7$, the evaluation of the AUC metrics decreases monotonically (AUC 0.66 at $\delta = 0.5$). At values $\delta > 1.0$, we observe a total reduction of the AUC values. Specifically, models trained for $\delta = 1.1$, and $\delta = 1.2$ show low AUCs as 0.58 and 0.56 but the lowest value (AUC value 0.51) is reached at $\delta = 1.5$. The evaluation of the loss function (when the reference data are propagated through the CNN models) confirms the scenario expressed above [see Fig. 4(b)]. Indeed, the global minimum value is achieved at $\delta = 1.0$ and a relative minimum is also present at $\delta = 0.8$. When δ is decreased or augmented toward the bound values $\delta = 0.5$ and $\delta = 1.5$, respectively, we can observe the loss function tends to reach higher values. In particular, an increase or decrease in the AUC along the δ domain is followed by a decrease or an increase in the loss function.

The application of the combined SMOE Scale on the mean instances (namely, the instances obtained by averaging all samples of the test data of the reference case) can help identify which patterns in the data are captured by the CNN to generate (accurate or degraded) ONI predictions. The reason for analyzing the mean instance is that it represents the main patterns in the feature time series; the interpretation of the saliency maps of all instances would turn out to be really impractical. The mean instances of both the events El Niño and La Niña are represented in Figs. 5(a) and 6(a), respectively. Therefore, after propagating the mean instance through the trained CNN model, we get the activated feature maps and compute the saliency map by means of the SMOE scale method. Next, we individualize the regions (months) of the saliency maps achieving the highest values; at the same regions of the mean instance, we can identify those time-series patterns that are mainly captured by the CNN model.

Taking into account the event El Niño, the saliency map of $\delta = 1$ reference case [green line in Fig. 5(a)] shows two peaks with an intensity of 0.6 and 0.9 around months 18 and 36, respectively (note that the instances are 48 months long and that the lead time is 9 months). These two regions turn out to be the most salient along the whole domain of the mean instance. At month 18, we find a peak in the thermocline depth [Fig. 5(a), green line] and a trough in both sea surface temperature [Fig. 5(a), orange line] and wind speed [Fig. 5(a), indigo line]. Conversely, in the neighborhood of month 36, we find the thermocline is descending toward a trough, while both sea surface temperature and wind speed are reaching a peak value. Both these two combinations of patterns represent the main characteristic that mostly defines the event El Niño according to the recognition activity of the CNN model.

Likewise, we can find some similar results for the event class La Niña. The spatial locations, where the saliency map [Fig. 6(b), green line] achieves values close to unity corresponding to one interval domain (months 0–7) of the mean instance [see Fig. 6(a)] where the thermocline depth attains a peak while both

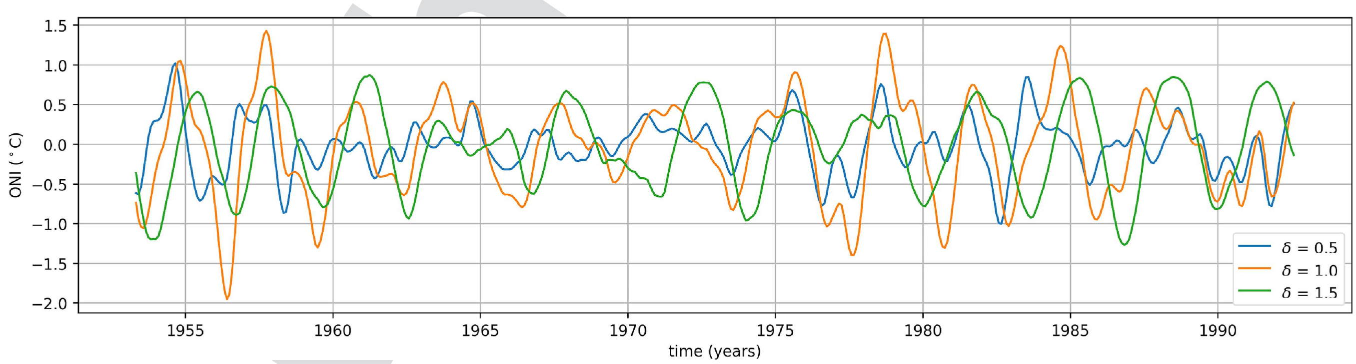

FIG. 3. Several time series of ONI calculated from ZC model simulations using δ parameter values of 0.5, 1.0, and 1.5, respectively.

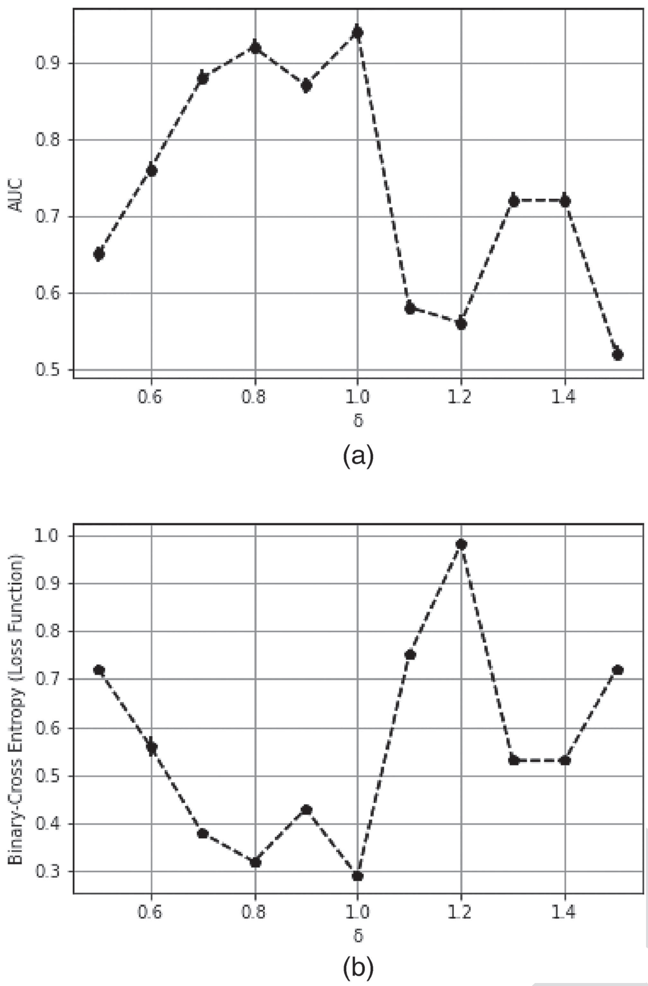

FIG. 4. The AUC score (a) and the loss function (b) as a function of the equatorial wave speed δ. Each point represents the mean AUC over five different folds; error bars are evaluated via the standard error mean.

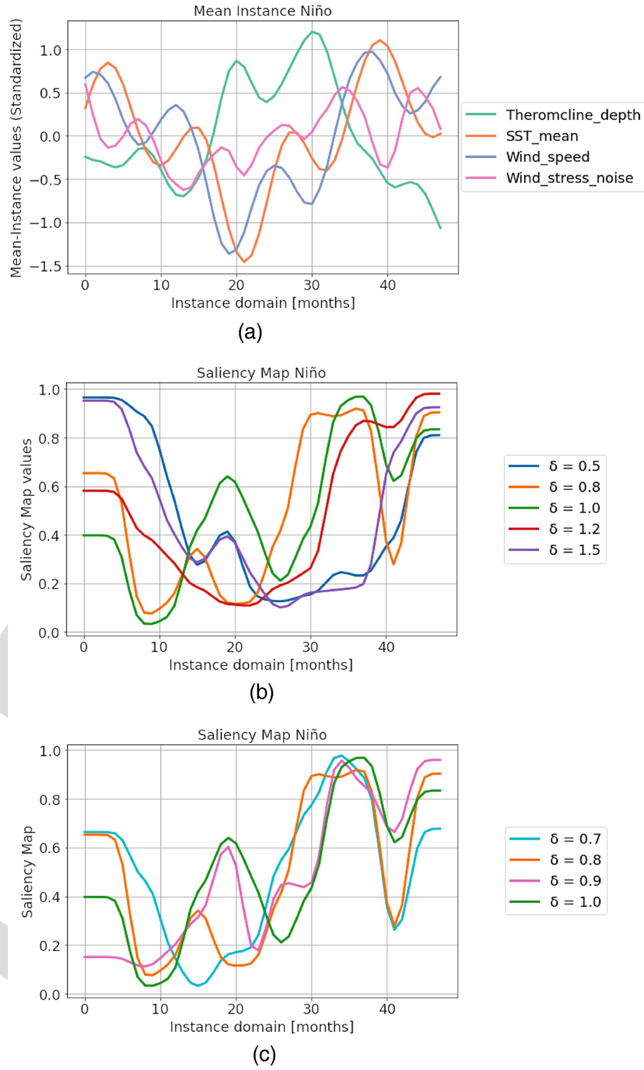

FIG. 5. The mean instance considering all the El Niño event instances in the test data (reference case) (a). Saliency maps of CNN models (b) and (c) trained with the wave distorted data (variation of δ) considering the cases (b) $\delta = 0.5, 0.8, 1.0, 1.2, 1.5$ and (c) $\delta = 0.7, 0.8, 0.9, 1.0$.

the sea surface temperature and the wind speed descend toward a trough.

When considering other distorted cases, such as $\delta = 0.5$ and $\delta = 0.8$ (where waves are propagating faster than the reference case), the combination of peaks in the thermocline and troughs in the sea surface temperature (and vice versa) still represents those relevant time-series patterns that the CNN model captures during the learning phase. If we focus our attention on the event El Niño, the saliency map of case $\delta = 0.8$ [see Fig. 5(b), orange line] shows at months 30–38 a broad region with intensity larger than 0.8, whereas the saliency map of case $\delta = 0.5$ [Fig. 5(b), blue line] shows intensities close to unity in the region 0–10 months. In the first case, we find that the high saliency region corresponds to a peak in the thermocline and a trough in both the sea surface temperature and the wind speed, while in the latter case, we find the thermocline depth shows a soft minimum value as opposed to the high-valued peak in the sea surface temperature and wind speed.

Similar results are also obtained for the event La Niña. The saliency map of case $\delta = 0.5$ [see Fig. 6(b), blue line] shows at months 35–40 a saliency region with intensity higher than 0.85. In Fig. 6(a), we see that this high saliency region corresponds to a peak in the thermocline depth and a trough in both the sea surface temperature and the wind stress. The saliency map of case $\delta = 0.8$ [see Fig. 6(b), orange line] reveals a broad high-valued peak (maximum intensity 0.81) around month 18 and a flat salient region (intensity close to 0.9) at months 40–48. By looking at those temporal regions

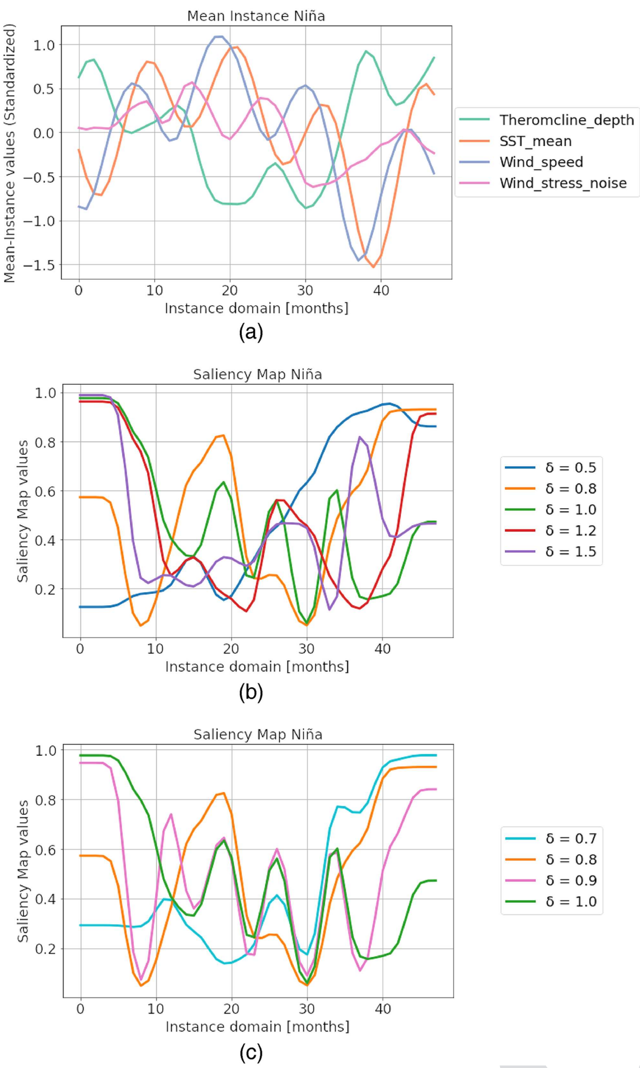

FIG. 6. The mean instance considering all the event La Niña instances in the test data (reference case) (a). Saliency maps of CNN models (b) and (c) trained with the wave distorted data (variation of δ) considering the cases (b) $\delta = 0.5, 0.8, 1.0, 1.2, 1.5$ and (c) $\delta = 0.7, 0.8, 0.9, 1.0$.

in Fig. 6(a), we find a peak in both the sea surface temperature and the wind stress together with a deep trough in the thermocline depth around month 18, whereas the domain months 40–48 show a soft trough in the thermocline and a prominent peak in both sea surface temperature and wind speed.

A deeper insight into the region $\delta = 0.7$–1.0 (where the AUC takes the highest values) reveals that all CNN models tend to capture a specific type of time-series patterns when they have to deal with the recognition of the event El Niño. For all cases considered in this interval, the saliency maps indicate as interesting the region months 32–38 [see Fig. 5(c)], where the values attained are larger than 0.8. Thus, we find that the results previously discussed for both cases $\delta = 0.8$ and $\delta = 1.0$ (where we discussed the behavior of the mean instance around month 36) are still valid even when we consider both cases $\delta = 0.7$ and $\delta = 0.9$. Interestingly, the case $\delta = 0.9$ shows a recognition activity that is similar to that of the model trained under the reference case because the saliency maps appear to be partially overlapped [see both the pink and the green line of Fig. 5(c) at months 32–38]. Moreover, the saliency map of case $\delta = 0.9$ points out some other aforementioned details of interest, e.g., those corresponding to the peak with intensity 0.6 at month 18, as shown by the pink line in Fig. 5(c).

For La Niña event, instead, the saliency map of case $\delta = 0.7$ [Fig. 6(c), cyan line] presents some analogies with case $\delta = 0.8$ [Fig. 6(c), orange line], individualizing one highly salient region at months 40–48 with an intensity around. Similarly to case $\delta = 1.0$ [Fig. 6(c), green line], the saliency map of case $\delta = 0.9$ [Fig. 6(c), pink line] individualizes a salient region in proximity of the left edge of the instance domain (months 0–5) with an intensity around to 0.95. It is interesting to note that both saliency maps of cases $\delta = 0.9$ and $\delta = 1.0$ are overlapped at the middle region (months 15–35); both two CNN models show a similar approach to capturing some low relevant features to identify the event La Niña.

For the cases $\delta = 1.2$ and $\delta = 1.5$ (where waves are propagating slower), the saliency maps [Fig. 5(b), red and purple line] reveal that the region around month 18 is no longer salient as in the reference case for El Niño event. For case $\delta = 1.2$, we find a salient region at months 36–48, where the saliency map takes values larger than 0.8. This corresponds to the presence of one broad peak in both the sea surface temperature and the wind speed with a less important contribution (than in the reference case) in the thermocline depth located at months 32–48. For case $\delta = 1.5$, we can observe that the saliency map is similar and almost completely overlapped with that of case $\delta = 0.5$; in this case, the analysis of the most salient time-series patterns will lead to some results that have already been discussed for the case $\delta = 0.5$.

When considering the event La Niña, the saliency map of case $\delta = 1.2$ [see Fig. 6(b), red line] attains values with an intensity close to unity at months 0–10. This temporal domain is characterized by the opposite feature, i.e., a broad peak in the thermocline depth and a trough in the sea surface temperature, located at months 0–10. The same feature can be also found for the case $\delta = 1.5$. Indeed, the saliency map [see Fig. 6(b), purple line] shows high saliency regions at either months 0–10 (intensity values close to unity) and month 36 (a peak with a maximum of 0.81). In particular, at month 36, the sea surface temperature and the wind speed reach a deeper trough with respect to that of region months 0–10.

C. Equatorial wave dynamics: Filtering of the instances via Butterworth digital filter

The application of a bandpass filter on all the instances included in the test dataset (reference case data) reveals that the propagation of one specific frequency band through the CNN models can retrieve most of the AUC scores obtained with the non-filtered data, as shown in Fig. 7. In specific, the model trained under the reference case turns out to be very sensitive to the frequency band corresponding to periods 8–16 months, where the AUC is equal to 0.80 [Fig. 7(a), green line]. On the contrary, the complete

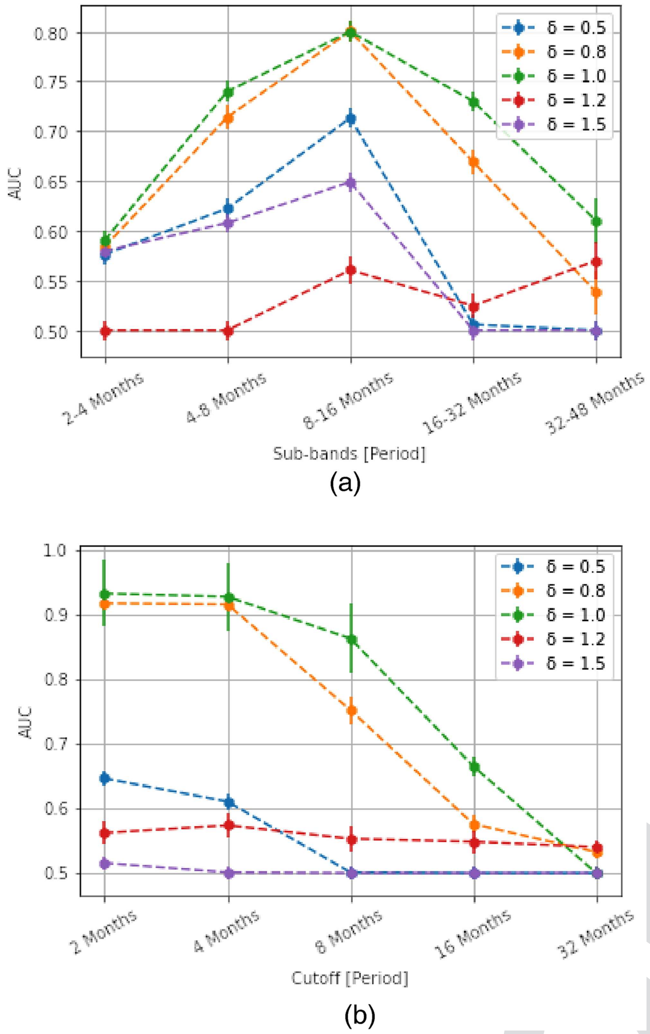

FIG. 7. The AUC score for different values of δ for the event El Niño as a function of (a) the bandpass frequency range and (b) the cut-off frequency, obtained by filtering the data by (a) bandpass Butterworth digital filter and (b) a low-pass Butterworth digital filter.

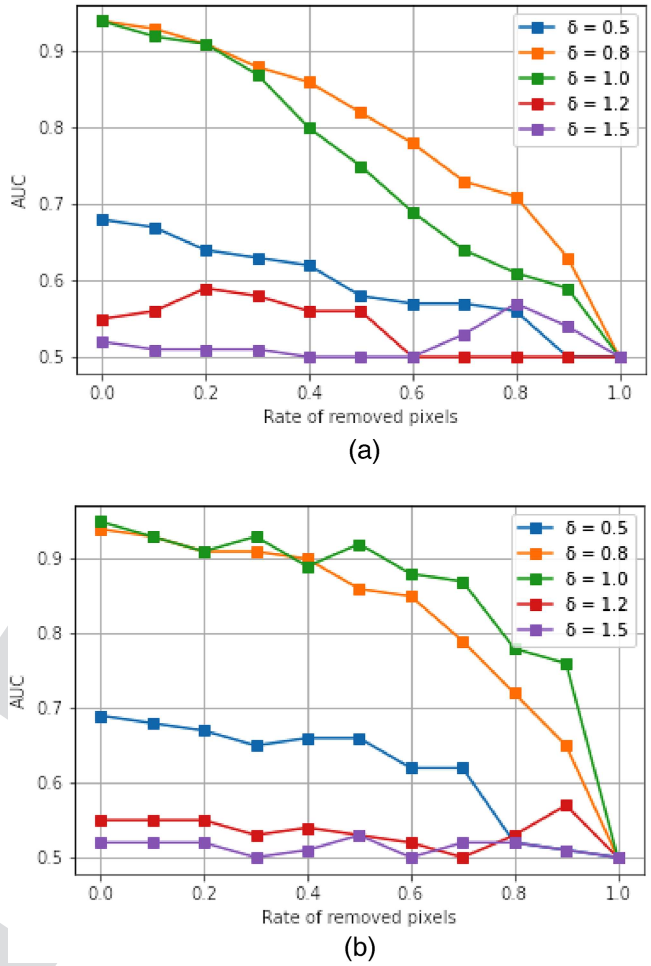

FIG. 8. Evaluation of AUC when the ROAR method (a) or the replacing at random strategy is applied (b); on the x axis, the ratio of pixels is replaced and on the y axis the AUC value.

degradation of AUC scores is attained when propagating lower and higher frequency bands, e.g., both intervals 16–32 months and 2–4 months, where the AUC value is equal to 0.61 and 0.58, respectively. Similar results can be found for other cases taken under consideration, as the case $\delta = 0.5$ and $\delta = 0.8$ [Fig. 7(a), blue and orange lines]. For both these cases, the band 8–16 months turns out to be the most predictive one with a net degradation of AUC score as soon as slower frequency bands are considered.

In particular, case $\delta = 0.8$ still shows some analogies with the reference case; the frequency band 8–16 months is still the most predictive with an AUC score of 0.80, and net degradation occurs at either lower or higher frequency bands. Such a result offers further details in interpreting the saliency maps, i.e., the CNN models tend to capture oscillating trends with specific carrier frequencies within the low-medium band of frequencies. It is important to highlight that the presence of details on a shorter frequency scale (i.e., period 16–32 months) is still fundamental and needed to allow the CNN to make an accurate classification of the ENSO events. The smoothing of the sample instances with a low-pass filter [Fig. 7(b)] reveals the instances tend to substantially lose many of their discriminating patterns at cutoff frequencies as 8 or 16 months. For example, in the cases $\delta = 0.8$ and $\delta = 1.0$ [Fig. 7(b), orange and green lines], we can observe a decrease in the predictive power with degradation of 0.1 AUC at 8 months and 0.3 AUC at 16 months. Hence, medium-low frequency patterns (4–8 months) as those contained in the thermocline depth or in the wind-noise time series can play an important role in the detection of the events.

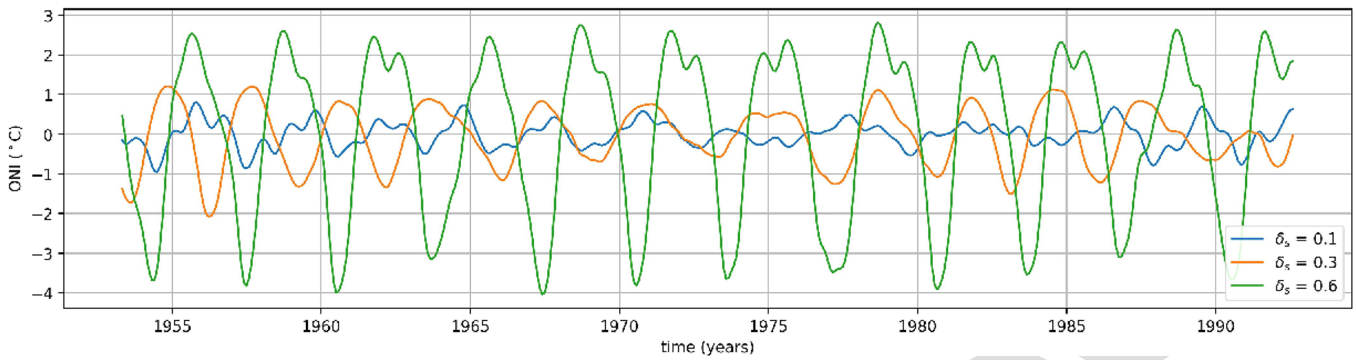

FIG. 9. Several time series of ONI calculated from ZC model simulations using δ_s parameter values of 0.1, 0.3, and 0.6 using $\mu = 2.7$.

D. Equatorial wave dynamics: ROAR

To ensure the correct implementation of the combined SMOE Scale and guarantee the validity of the results obtained, we used (and adapted to this analysis) the metrics ROAR (Remove and Retain) introduced in Mundhenk *et al.* (2019). The replacement in the validation sets of an increasing amount of salient spatial locations with zero-valued pixels rapidly deteriorates the predictive characteristics of the data; as shown in Fig. 8. It is important to remember that the CNN models do not make use of any bias term neither in the convolutional layers nor in the dense layers. Accordingly, the CNN model considers the zero-valued patterns as absolutely non-informative, i.e., the propagation of such a pattern through the CNN is designed to prevent the activation of any stimulus along the hidden layers. In Fig. 8(a), we can observe that the removal of the top 50% salient pixels via ROAR (actually 24) guarantees a considerable decrease in the AUC; under the reference case model, the AUC scores present a loss equal to 0.20. Contrary to this, when randomly replacing the 50% pixels with zero-valued pixels, we can still observe a slighter decrease in the AUC curve under the reference, i.e., a loss equal to 0.03 [see Fig. 8(b)]. Likewise, similar results can be found when even considering all the other distorted physics cases.

E. Upwelling feedback: Saliency maps

We next consider the distortion of the model data due to a wrong representation of the upwelling feedback, represented by the parameter δ_s in the ZC model. Figure 9 shows that the ONI's amplitude increases (decreases) for larger (smaller) values of δ_s. This behavior is expected because the upwelling feedback is a positive one, enhancing the existing sea surface temperature anomaly further and consequently increasing the amplitude of the ONI. The AUC score vs δ_s curve [Fig. 10(a)] reveals that a particular tuning of the parameter δ_s strongly affects the accuracy of the CNN models when trained with distorted data. By construction, the AUC score attains the highest score at the reference value $\delta_s = 0.3$ (AUC 0.94). For $\delta_s < 0.3$ the profile of the curve suggests a net degradation in the AUC scores with the lowest score attained at $\delta_s = 0.15$ (AUC 0.5), whereas at $\delta_s > 0.3$ the AUC scores remain stable, but still attain values lower than 0.7. The profile of the AUC has a plateau at values of 0.6 as δ_s goes toward the boundary value $\delta_s = 0.6$. The evaluation of the loss function [Fig. 10(b)] as a function of the parameter δ_s confirms the results obtained above. At $\delta_s = 0.3$, the global minimum is achieved, and the net degradation occurring at lower and higher $\delta_s = 0.3$ are still present; the loss function increases monotonically in both cases. Similarly to the analysis provided for the distortion of the δ parameter, we next consider the mean instances (of the test data of the reference case) and their saliency maps [Figs. 11(a) and 12(a)].

For the event El Niño, we can observe that different regions of saliency can be associated to different variations of δ_s, i.e., for $\delta_s < 0.3$ the saliency maps [Fig. 11(b), blue and orange lines] indicate the left part of the instance as the most predictive, while for $\delta_s > 0.3$ the right part [Fig. 11(b), red and purple lines]. In particular, the saliency map of cases $\delta_s = 0.10$ and $\delta_s = 0.25$ [Fig. 11(b), blue and orange lines] turns out to be very salient at 0–8 months, with intensity above 0.8. In that region, the mean instance presents a peak occurring in both the sea surface temperature and the wind speed time-series features. On the contrary, for cases $\delta_s = 0.45$ and $\delta_s = 0.60$, the saliency maps [Fig. 11(b), red and purple lines] achieve intensities larger than 0.8 around 32–48 months and capture one single broad oscillating peak in both the sea surface temperature and the wind speed time-series features.

For the event La Niña, we refer to Fig. 12. In particular, the saliency maps of cases $\delta_s = 0.10$ and $\delta_s = 0.25$ [Fig. 12(b), blue and orange lines] present intensities larger than 0.8 at 42–48 months. It is interesting to observe that the saliency map of case $\delta_s = 0.25$ presents a plateau around 32–48 months; in opposition to the event El Niño, the CNN here captures a deep trough in the sea surface temperature time-series feature.

F. Upwelling feedback: Filtering of the instances via Butterworth digital filter

The application of bandpass and low-pass filters on the sample instances brings to light a result similar to the analysis done for the parameter δ, as shown in Fig. 13. When applying a bandpass filter with bandwidth 8–16 months, the case $\delta_s = 0.25$ [Fig. 13(a), orange line] can partially retrieve the original prediction with AUC 0.70, whereas for other cases such as $\delta_s = 0.6$ [Fig. 13(a), purple line] the

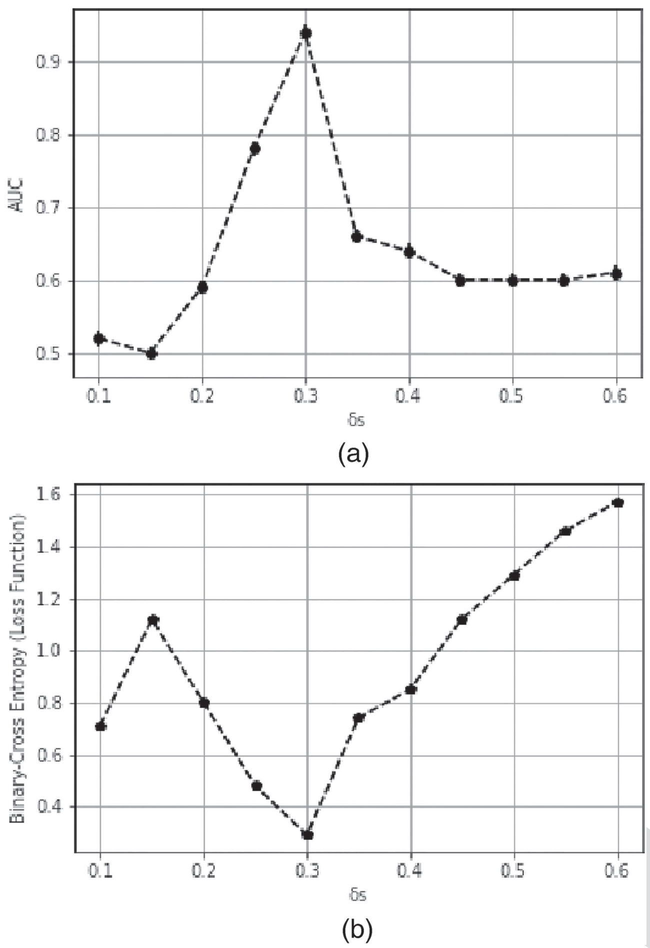

FIG. 10. The AUC score (a) and the loss function (b) as a function of the upwelling feedback parameter δ_s. Each point represents the mean AUC over five different folds; error bars are evaluated via the standard error mean.

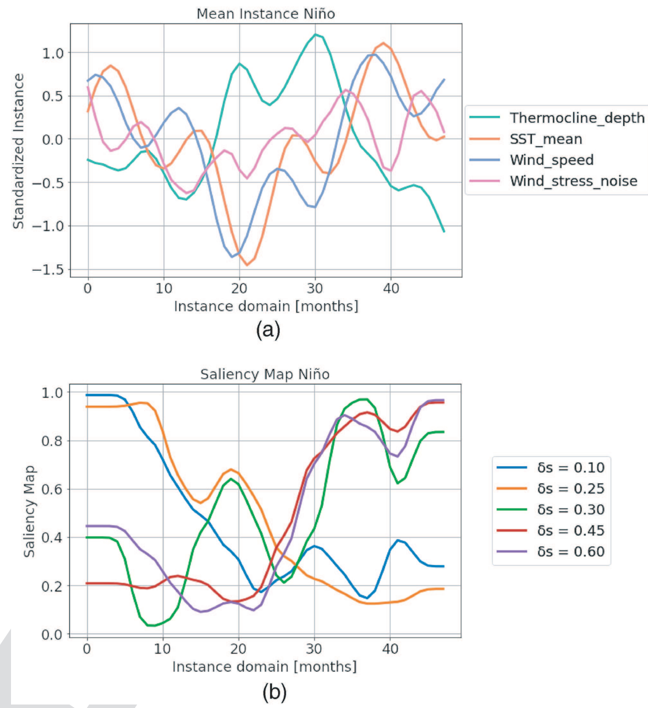

FIG. 11. The mean instance (a) of all the El Niño event instances in the test data (reference case). Saliency maps of CNN models (b) trained with the upwelling distorted data (variation of δ_s). In specific, cases $\delta_s = 0.10, 0.25, 0.30, 0.45, 0.60$ are considered.

original prediction can be retrieved only by oscillations lying within the frequency band corresponding to 32–48 months. The smoothing of the instances via low-pass filter [Fig. 13(b)] shows that the removal of high-frequency patterns oversimplifies the data; and so, the classification task cannot be solved by the information contained in the low-frequency data only.

As confirmed by the filtering of the instances, the frequency bands 4–8 months and 8–16 months represent the main frequency bands in the reference case ($\delta_s = 0.3$). Capturing one of these two can retrieve a considerable amount of skill. The case $\delta_s = 0.25$ focuses a large amount of relevant patterns mainly in the frequency band 8–16 months. The filtering with a low-pass digital filters also reveals that a cut-off frequency of 16 months can reduce the AUC in both cases, but a cut-off frequency of 8 months leads to a degradation for the reference case only. In the latter scenario, we register a loss of 0.1 AUC, i.e., a degradation on the same order of magnitude as when testing the reference case data and the data of case $\delta_s = 0.25$. Hence, this example shows how a manipulation in the intrinsic characteristic of the instances can lead to a reduction and oversimplification of the instances, i.e., the distortion of the periodicity of data provokes a reduction or missing of some patterns that are fundamental in the classification of the reference case data.

G. Comparison of CNN and GDNN

To provide a comparison, we also applied the distorted physics approach in the Gaussian Density Neural Network (GDNN) as used in Petersik and Dijkstra (2020). The Gaussian density terminology refers to the network's purpose of predicting a Gaussian distribution by producing both a mean and standard deviation as output. The variable to be predicted (or target variable) is also the ONI at a (lead) time in the future. The features used in the GDNN are described by Petersik and Dijkstra (2020): ONI, network graph connectivity metric c_2, adjusted Hamming distance $\mathcal{H}^*$ (measure of change in the network graph) and a seasonal cycle (SC) in the form of a cosine. The warm water volume (WWV, volume of water above the 20 °C thermocline) is not available in the output of ZC model, and, therefore, the thermocline depth itself was used here. All feature datasets are normalized before training.

Training the GDNN consists of a number of ensemble members that are trained in parallel. Each of the members is trained for 100 iterations over 500 epochs with a batch size of 100. The training starts with a random selection of hyperparameters within

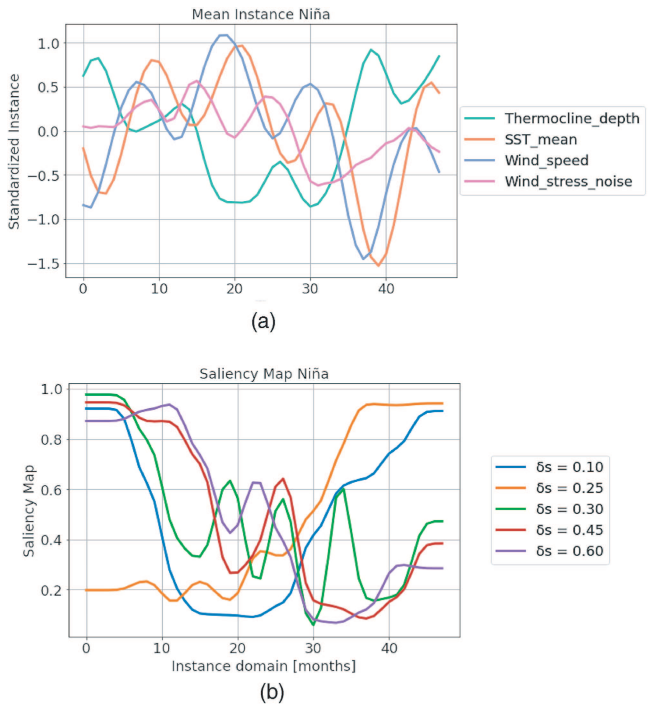

FIG. 12. The mean instance (a) of all the La Niña event instances in the test data (reference case). Saliency maps of CNN models (b) trained with the upwelling distorted data (variation of δ_s). In specific, cases $\delta_s = 0.10, 0.25, 0.30, 0.45, 0.60$ are considered.

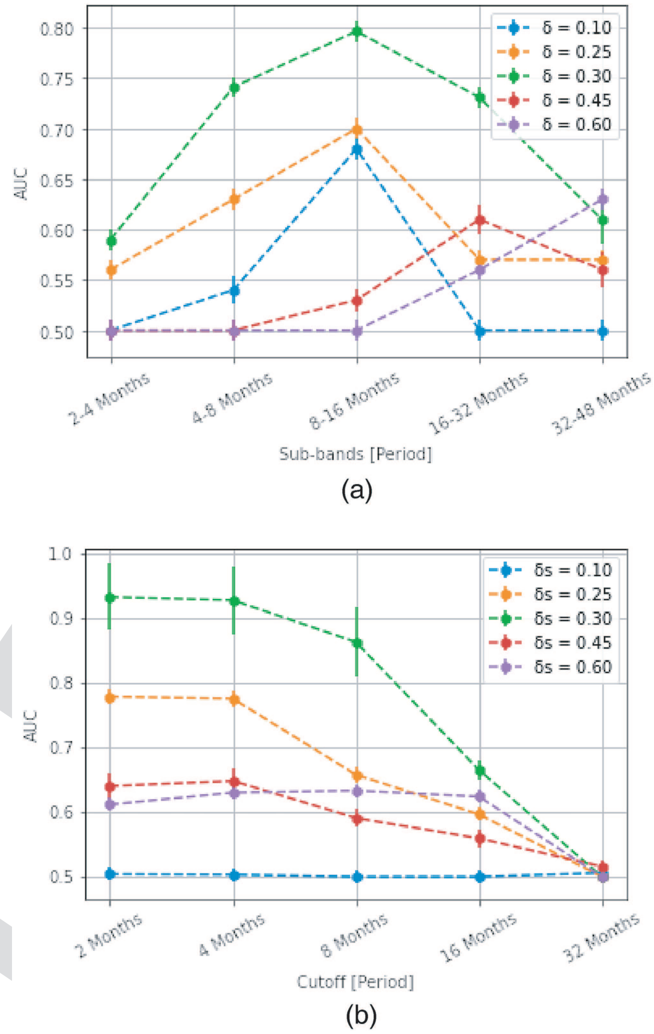

FIG. 13. The AUC score for different values of δ_s for the event El Niño as a function of (a) the bandpass frequency range and (b) the cut-off frequency, obtained by filtering the data by (a) bandpass Butterworth digital filter and (b) a low-pass Butterworth digital filter.

bounds defined by the user and is then optimized using the ADAM algorithm (Kingma and Ba, 2014) with a user specified learning rate, dropout, and Gaussian noise. The resulting ensemble members each predict a mean and standard deviation of the target variable and these predictions are then averaged over the ensemble for the final prediction. Again, the lead time is 9 months in the result below.

We use two different measures for the performance of the GDNN: the RMSE and the Pearson correlation; also, the loss function is shown (see Fig. 14). Different simulations give different networks and give different performance values. The GDNN's, when trained on distorted physics data, still perform consistently when varying δ or δ_s. However, a change in the ONI's amplitude in the training data (such as for higher than reference δ_s) is poorly corrected for, leading to a large overestimation of the predicted variable [e.g., see $\delta_s = 0.40$ in Fig. 14(a)]. The model only tolerates a difference in amplitude between test and training dataset ONI if only a small distortion of the variable is used (e.g., $\delta_s = 0.35$). The ability to compensate for the period but not the amplitude is explained by the relatively simple architecture of the GDNN. Whereas the former only requires a scalar addition to the input, the latter would require some linear combination of (co)sines to be learned by the neural network.

The attempt of comparing the capability of both CNN and GDNN in detecting El Niño events is made complicated by the intrinsic design of both models. Although both models are trained to solve the same problem, we have to take into account that the CNN model is a binary classifier, while the GDNN is designed to solve regression problems. In addition, the fact that both models optimize the same loss function does not ensure a relation or a similarity about what the two models learn during the training phase can be found. The two models could focus on capturing totally different features of data, because the outputs of the two models represent two different probabilities, i.e., the CNN estimates the probability of the event itself, whereas the GDNN estimates the probability distribution of the ONI index. However, the ENSO events are based on the behavior of the ONI index and we can exploit this fact to make the outputs of the GDNN more close to those of the CNN. After training the GDNN, we can use the estimation on the Gaussian density

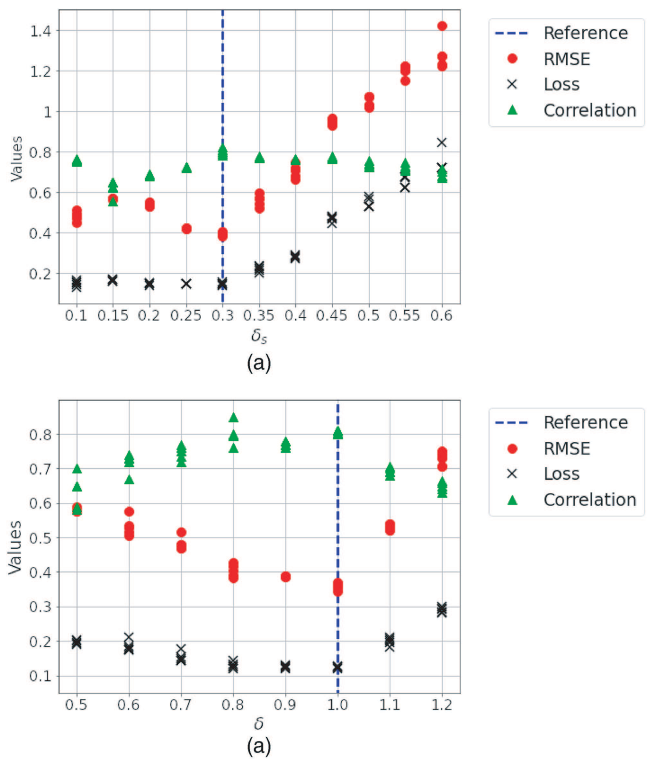

FIG. 14. Performance of the GDNN when trained on distorted ZC model data using several values of (a) δ_s and (b) δ.

to estimate the probability of El Niño events, i.e., the probability that the absolute value of ONI index is greater than 0.5 °C. Thereafter, we can use the AUC metric to compare the performance of the two models.

As we can see in Fig. 15, the GDNN model appears to be less accurate than the CNN model. The reference case data show a lower AUC [compare to Fig. 3(a)] and we can observe a general reduction of 0.1 AUC with respect to the results obtained with the CNN model. When feeding the GDNN model with ZC data with a different tuning of parameters δ, we can observe that GDNN tends to be more degraded at $\delta < 1$, then the CNN model [compare to Fig. 8(a)]; in fact, the AUC can lose up to 0.21 with respect to the reference case. Note that the same tuning of parameter δ would reveal a plateau in the AUC score whose values are much closer to that one attained in the reference case. When considering the distortion of parameter δ_s we can still appreciate a degradation at values lower than 0.3. However, the decrease in the AUC scores appears milder (~ 0.1) with respect to that shown for the CNN model. On the contrary, as $\delta_s > 0.3$, there is a significant reduction in the AUC scores; with respect to the reference case, the AUC scores can now be reduced up to 0.2.

IV. SUMMARY AND DISCUSSION

This work was strongly motivated to understand the high skill in ENSO prediction obtained with the CNN approach in

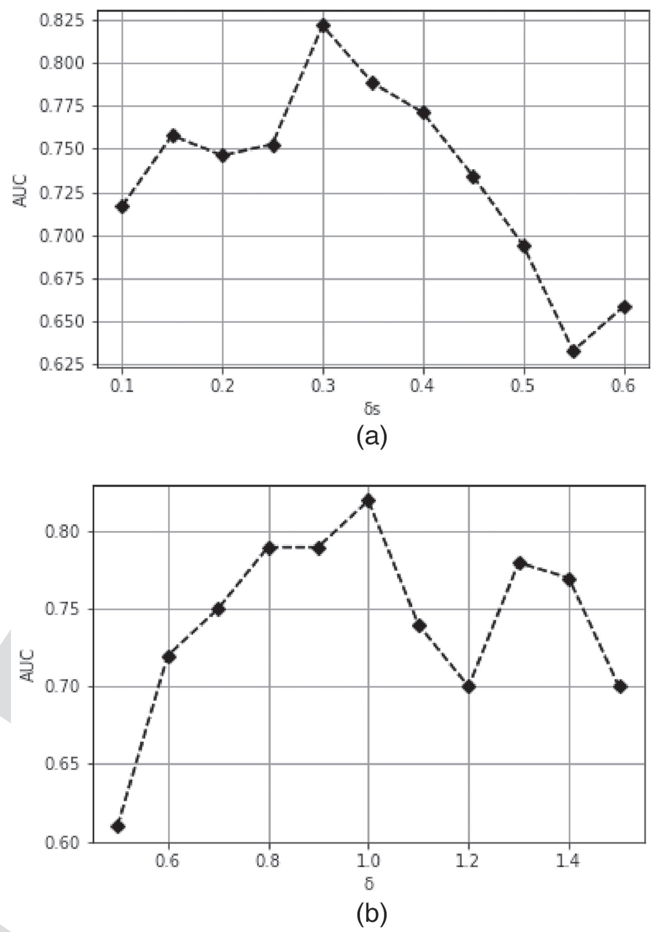

FIG. 15. AUC metric for the GDNN when considered as a classifier for both the wave distorted case (a) and the upwelling distorted case (b). On the x axis, the values of ZC parameters (δ, δ_s) and on the y axis the AUC score.

Ham *et al.* (2019), in particular, at long lead times. Although heat maps were presented in Ham *et al.* (2019), their analysis does not connect immediately to the detailed processes of ENSO dynamics, which is also difficult because of the wide range of data they used. In this paper, we introduced distorted physics simulations with the well-known Zebiak–Cane (ZC) model (Zebiak and Cane, 1987) to determine how a CNN can perform on real data when trained on data from "wrong" model simulations.

The behavior of the ZC model can be elegantly described by a delay-differential equation (Suarez and Schopf, 1988; and Jin, 1997)

$$\frac{dT(t)}{dt} = aT(t) - bT(t-d) - cT^3(t) \qquad (2)$$

for the eastern Pacific temperature T as a function of time t. Here, the constant a indicates the strength of the positive feedbacks, b that of the delayed negative feedback (with a delay d due to equatorial wave dynamics), and c measures the strength of the nonlinear equilibration.

By distorting the δ parameter in the ZC model, we modify the delay d in (2) and, hence, mostly the adjustment processes in the equatorial Pacific. When the equatorial wave speeds are distorted, there is an asymmetry in the skill of the CNN. For faster waves $\delta < 1$, the performance remains good, whereas for $\delta > 1$ (slower waves), it deteriorates. For example, in case $\delta = 1.2$, the El Niño event appears to be mainly constituted by slower oscillations, even though the behavior of the large-scale thermocline depth and sea surface temperature is similar to the reference case. However, the loss of details on shorter time scales leads the model to still reasonably solve the classification task.

By distorting the parameter δ_s, we basically modify the feedback parameter a in (2) and, hence, the amplitude of the El Niño events. However, also the stability properties of the background climate state are changed as seen through the shift in the Hopf bifurcation with δ_s (van der Vaart *et al.*, 2000). For increasing δ_s and constant μ (as is done here), the background destabilizes as can also be seen in Fig. 9. The case $\delta_s = 0.1$ (reference case $\delta_s = 0.3$) offers a clear example about how the manipulation in the upwelling feedback can degrade the AUC, i.e., the distortion of the patterns in the data leads to a misplacement and misalignment and reduce the capability of the network in capturing the right patterns at the right (temporal) location. For other cases (e.g., $\delta_s = 0.25$, $\delta_s = 0.45$, and $\delta_s = 0.6$), the skill of the CNN predictions is reduced less, because the right combination of peaks and valleys in the time series are present. Indeed, the absence of oscillating terms located at the frequency band 4–8 months does not allow the CNN to capture all the relevant patterns but only a part of them.

The results indicate that the accuracy of the classification of the El Niño and La Niño events for lead times of 9 months using a CNN approach is strongly related to the capability of the CNN to capture the wave adjustment and feedback processes. The exact combination of specific patterns like peaks and valleys occurring at specific regions of the time domain of all features is essential to generate skill in the CNN predictions. The distorted physics approach can be very useful to look at how a CCN based prediction scheme can represent additional processes. For example, it is well known that connections between the Indian-Pacific (Izumo *et al.*, 2010) and Atlantic-Pacific (Ham *et al.*, 2013) and extratropical-tropical connections (Zhao and Di Lorenzo, 2020) are important for the skill of ENSO predictions. The latter interactions have been described as ocean-atmosphere meridional modes and can influence ENSO and tropical variability on decadal time scales from both hemispheres independently (Amaya, 2019). Also, the effect of climate change on ENSO prediction skills, and how a CNN would capture this, is an interesting future line of work. However, one cannot use the Zebiak–Cane model for such studies and needs to do such distorted physics simulations with more sophisticated global climate models.

ACKNOWLEDGMENTS

The work by H.D. was sponsored by the Netherlands Science Foundation (NWO) through Project No. OCENW.M20.277.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

All authors contributed to the design of this study. Results were mainly obtained by G.L. and I.G. The paper was jointly written with contributions from all authors.

G. Lancia: Data curation (supporting); Formal analysis (supporting); Methodology (lead); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **I. J. Goede:** Data curation (lead); Formal analysis (equal); Investigation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **C. Spitoni:** Formal analysis (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **H. Dijkstra:** Formal analysis (lead); Methodology (equal); Supervision (lead); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in github at https://github.com/glancia93/Physics-captured-by-data-based-methods-in-El-Nino-prediction_PyCODE, Ref. ■

REFERENCES

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., and Kim, B., "Sanity checks for saliency maps," arXiv:1810.03292 (2018).

Amaya, D. J., "The Pacific meridional mode and ENSO: A review," Curr. Clim. Change Rep. **5**, 296–307 (2019).

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion **58**, 82–115 (2020).

Balmaseda, M. A., Davey, M. K., and Anderson, D. L. T., "Decadal and seasonal dependence of ENSO prediction skill," J. Clim. **8**, 2705–2715 (1995).

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., and DeWitt, D. G., "Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing?," Bull. Am. Meteorol. Soc. **93**, 631–651 (2012).

Barnston, A. G., Tippett, M. K., Ranganathan, M., and L'Heureux, M. L., "Deterministic skill of ENSO predictions from the North American multimodel ensemble," Clim. Dyn. **53**, 7215–7234 (2019).

Butterworth, S., "On the theory of filter amplifiers," Wireless Eng. **7**, 536–541 (1930).

Chen, D. and Cane, M. A., "El Niño prediction and predictability," J. Comput. Phys. **227**, 3625–3640 (2008).

Diaz, H. F., Hoerling, M. P., and Eischeid, J. K., "ENSO variability, teleconnections and climate change," Int. J. Climatol. **21**, 1845–1862 (2001).

Dijkstra, H. A., *Atmospheric and Oceanographic Sciences Library* (Springer, 2005), Vol. 28, p. 480.

Dijkstra, H. A. and Neelin, J., "Coupled ocean-atmosphere models and the tropical climatology. II: Why the cold tongue is in the east," J. Clim. **8**, 1343–1359 (1995).

Dijkstra, H. A., Petersik, P., Hernández-García, E., and López, C., "The application of machine learning techniques to improve El Niño prediction skill," Front. Phys. **7**, 153 (2019).

Duan, W., Liu, X., Zhu, K., and Mu, M., "Exploring the initial errors that cause a significant 'spring predictability barrier' for El Niño events," J. Geophys. Res. **114**, C04022 (2009).

Fedorov, A., Harper, S., Philander, S., Winter, B., and Wittenberg, A., "How predictable is El Niño?," Bull. Am. Meteorol. Soc. **84**, 911–919 (2003).

Feng, Q. Y. and Dijkstra, H. A., "Climate network stability measures of El Niño variability," Chaos **27**, 035801 (2017).

Gill, A., "Some simple solutions for heat-induced tropical circulation," Q. J. R. Meteor. Soc. **106**, 447–462 (1980).

Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning* (MIT Press, 2016).

Ham, Y.-G., Kim, J.-H., and Luo, J.-J., "Deep learning for multi-year ENSO forecasts," Nature **573**, 568–572 (2019).

Ham, Y.-G., Kug, J.-S., Park, J.-Y., and Jin, F.-F., "Sea surface temperature in the north tropical Atlantic as a trigger for El Niño/southern oscillation events," Nat. Geosci. **6**, 112–116 (2013).

Hamming, R. W., *Digital Filters* (Courier Corporation, 1998), pp. 561–569.

Hou, M., Duan, W., and Zhi, X., "Season-dependent predictability barrier for two types of El Niño revealed by an approach to data analysis for predictability," Clim. Dyn. **53**, 5561–5581 (2019).

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montegut, C., Behera, S. K., Luo, J.-J., Cravatte, S., Masson, S., and Yamagata, T., "Influence of the state of the Indian Ocean Dipole on the following year's El Niño," Nat. Geosci. **3**, 168–172 (2010).

Jin, F.-F., "An equatorial recharge paradigm for ENSO. I: Conceptual model," J. Atmos. Sci. **54**, 811–829 (1997).

Jin, F.-F., "An equatorial recharge paradigm for ENSO. II: A stripped-down coupled model," J. Atmos. Sci. **54**, 830–8847 (1997).

Jin, F.-F., Neelin, J., and Ghil, M., "El Niño on the devil's staircase: Annual subharmonic steps to chaos," Science **264**, 70–72 (1994).

Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," arXiv:1412.6980[cs] (2014).

Kug, J.-S., Jin, F.-F., and An, S.-I., "Two types of El Niño events: Cold tongue El Niño and warm pool El Niño," J. Clim. **22**, 1499–1515 (2009).

Lakshminarayanan, B., Pritzel, A., and Blundell, C., "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30, pp. 6402–6413.

Latif, M., "Dynamics of interdecadal variability in coupled ocean-atmosphere models," J. Clim. **11**, 602–624 (1998).

Latif, M. and Barnett, T. P., "Causes of decadal climate variability over the North Pacific and North America," Science **266**, 634–637 (1994).

L'Heureux, M. L., Takahashi, K., Watkins, A. B., Barnston, A. G., Becker, E. J., Di Liberto, T. E., Gamble, F., Gottschalck, J., Halpert, M. S., Huang, B., Mosquera-Vásquez, K., and Wittenberg, A. T., "Observing and predicting the 2015/16 El Niño," Bull. Am. Meteorol. Soc. **98**, 1363–1382 (2017).

Lian, T., Chen, D., Tang, Y., and Wu, Q., "Effects of westerly wind bursts on El Niño: A new perspective," Geophys. Res. Lett. **41**, 3522–3527, https://doi.org/10.1002/2014GL059989 (2014).

McPhaden, M. J., "Tropical Pacific Ocean heat content variations and ENSO persistence barriers," Geophys. Res. Lett. **30**, 2705–2709, https://doi.org/10.1029/2003GL016872 (2003).

McPhaden, M. J., Timmermann, A., Widlansky, M. J., Balmaseda, M. A., and Stockdale, T. N., "The curious case of the El Niño that never happened: A perspective from 40 years of progress in climate research and forecasting," Bull. Am. Meteorol. Soc. **96**, 1647–1665 (2015).

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., and Müller, K.-R., "Layer-wise relevance propagation: An overview," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (2019), pp. 193–209.

Mu, M., Sun, L., and Dijkstra, H. A., "The sensitivity and stability of the ocean's thermohaline circulation to finite amplitude perturbations," arXiv:0702083v1 [arXiv:physics] (2007), pp. 1–40.

Mundhenk, T. N., Chen, B. Y., and Friedland, G., "Efficient saliency maps for explainable AI," arXiv:1911.11293 (2019).

Neelin, J., "The slow sea surface temperature mode and the fast-wave limit: Analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model," J. Atmos. Sci. **48**, 584–606 (1991).

Neelin, J., Battisti, D. S., Hirst, A. C., Jin, F.-F., Wakata, Y., Yamagata, T., and Zebiak, S. E., "ENSO theory," J. Geophys. Res. **103**, 14261–14290, https://doi.org/10.1029/97JC03424 (1998).

Newman, M. and Sardeshmukh, P. D., "Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures?," Geophys. Res. Lett. **44**, 8520–8529, https://doi.org/10.1002/2017GL074088 (2017).

Petersik, P. J. and Dijkstra, H. A., "Probabilistic forecasting of El Niño using neural network models," Geophys. Res. Lett. **47**, 1–8, https://doi.org/10.1029/2019GL086423 (2020).

Preisendorfer, R. W., *Principal Component Analysis in Meteorology and Oceanography* (Elsevier, Amsterdam, 1988).

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., Van Den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E., "The NCEP climate forecast system version 2," J. Clim. **27**, 2185–2208 (2014).

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 618–626.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929–1958 (2014).

Suarez, M. and Schopf, P., "A delayed action oscillator for ENSO," J. Atmos. Sci. **45**, 3283–3287 (1988).

Tang, Y., Zhang, R.-H., Liu, T., Duan, W., Yang, D., Zheng, F., Ren, H., Lian, T., Gao, C., Chen, D., and Mu, M., "Progress in ENSO prediction and predictability study," Natl. Sci. Rev. **5**, 826–839 (2018).

Tian, B. and Duan, W., "Comparison of the initial errors most likely to cause a spring predictability barrier for two types of El Niño events," Clim. Dyn. **47**, 779–792 (2015).

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., Cobb, K., Lengaigne, M., McPhaden, M. J., Stuecker, M. F., Stein, K., Wittenberg, A. T., Yun, K.-S., Bayr, T., Chen, H.-C., Chikamoto, Y., Dewitte, B., Dommenget, D., Grothe, P., Guilyardi, E., Ham, Y.-G., Hayashi, M., Ineson, S., Kang, D., Kim, S., Kim, W., Lee, J.-Y., Li, T., Luo, J.-J., McGregor, S., Planton, Y., Power, S., Rashid, H., Ren, H.-L., Santoso, A., Takahashi, K., Todd, A., Wang, G., Wang, G., Xie, R., Yang, W.-H., Yeh, S.-W., Yoon, J., Zeller, E., and Zhang, X., "El Niño—Southern oscillation complexity," Nature **559**, 535–545 (2018).

Tziperman, E., Stone, L., Cane, M. A., and Jarosh, H., "El Niño chaos: Overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator," Science **264**, 72–74 (1994).

van der Vaart, P. C. F., Dijkstra, H. A., and Jin, F. F., "The Pacific cold tongue and the ENSO mode: A unified theory within the Zebiak-Cane model," J. Atmos. Sci. **57**, 967–988 (2000).

Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A. Y., "Temporal convolutional networks for the advance prediction of ENSO," Sci. Rep. **10**, 1–15 (2020).

Yu, Y., Mu, M., and Duan, W., "Does model parameter error cause a significant spring predictability barrier for El Niño events in the Zebiak-Cane model?," J. Clim. **25**, 1263–1277 (2012).

Zebiak, S. and Cane, M., "A model El Niño-southern oscillation," Mon. Weather Rev. **115**, 2262–2278 (1987).

Zhang, Z., Ren, B., and Zheng, J., "A unified complex index to characterize two types of ENSO simultaneously," Sci. Rep. **9**, 8373 (2019).

Zhao, Y. and Di Lorenzo, E., "The impacts of extra-tropical ENSO precursors on tropical Pacific decadal-scale variability," Sci. Rep. **10**, 1–12 (2020).

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A., "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 2921–2929.