OCEANOGRAPHY

Predicting Atlantic and Benguela Niño events with deep learning

Marie-Lou Bachèlery[1,2]*, Julien Brajard[3], Massimiliano Patacchiola[4], Serena Illig[5,6], Noel Keenlyside[1,3,7]

Atlantic and Benguela Niño events substantially affect the tropical Atlantic region, with far-reaching consequences on local marine ecosystems, African climates, and El Niño Southern Oscillation. While accurate forecasts of these events are invaluable, state-of-the-art dynamic forecasting systems have shown limited predictive capabilities. Thus, the extent to which the tropical Atlantic variability is predictable remains an open question. This study explores the potential of deep learning in this context. Using a simple convolutional neural network architecture, we show that Atlantic/Benguela Niños can be predicted up to 3 to 4 months ahead. Our model excels in forecasting peak-season events with remarkable accuracy extending lead time to 5 months. Detailed analysis reveals our model's ability to exploit known physical precursors, such as long-wave ocean dynamics, for accurate predictions of these events. This study challenges the perception that the tropical Atlantic is unpredictable and highlights deep learning's potential to advance our understanding and forecasting of critical climate events.

INTRODUCTION

Atlantic and Benguela Niño and Niña events are extreme climatic phenomena of the tropical Atlantic. Marked by anomalous interannual oceanic warming/cooling in the eastern equatorial Atlantic and southwestern coastal African regions (Fig. 1A), respectively, these events not only influence regional precipitation and climate patterns (*1–4*) but also play a substantial role in the coastal oxygen variability (*4–7*), fish habitat, and marine resource distribution/abundance. Atlantic Niño events also influence the Pacific El Niño Southern Oscillation (*8*, *9*). Therefore, operational forecasting of these events could be greatly beneficial to local ecosystem management and climate services globally.

It is particularly interesting and troubling that compared to the Pacific and despite notable efforts (*10*), there has been limited success in predicting these interannual events (*11–15*) to support management strategies. State-of-the-art dynamic forecasting systems even struggle to beat persistence—the most common benchmark—and to achieve statistical significance (Fig. 1, B and C). They completely failed in predicting the exceptionally strong 2021 events (*15*). This has led to a growing consensus that such events are inherently unpredictable (*11*, *16–19*, *20*, *21*).

Unlike the Pacific El Niño, the main challenge in forecasting the tropical Atlantic (*18*) is the notably weaker interannual variability compared to the seasonal cycle. In addition, systematic model biases in the tropical Atlantic affect atmospheric and oceanic dynamics (*20*, *22–24*) and may account for this lack of predictive capabilities. However, the difficulty in predicting Atlantic and Benguela Niño events goes beyond model bias (*11*, *20*, *25*), likely because of their complex dynamics, with nonlinear processes and interplays of oceanic and atmospheric connections (*26–29*). Although they peak in different seasons (Atlantic events in June-July and Benguela events in March-April-May), their developments are inextricably linked. Variations in the South Atlantic anticyclone induce anomalous coastal winds affecting the southwest African upwelling and initiating Benguela Niños/Niñas off the African coast (*28–31*). Simultaneously, the South Atlantic anticyclone variations trigger wind stress anomalies in the western part of the basin that prompt equatorial Kelvin waves and subsequent coastal trapped waves, influencing the equatorial (*32–35*) and coastal African ocean conditions (*36–41*). However, while the oceanic connection contributes to a major part of the Atlantic and Benguela events (*41–44*), as the events mature, strong air-sea interactions develop between the two events, either intensifying or attenuating them (*15*, *26*, *29*). Major nonlinearities arise from the Bjerknes feedback (*45*) controlling the Atlantic Niños, with the surface-subsurface coupling in the eastern equatorial basin strongly tied to the seasonal cold tongue development. Also, the ocean-atmosphere coupling is strongly modulated by the march of the Intertropical Convergence Zone (*46*). Capturing the complex mechanisms and nonlinear processes intrinsic to both the Atlantic and Benguela events in dynamic prediction systems is a challenge and may explain their lack of skill for these events.

While forecasting Atlantic and Benguela Niños/Niñas seems to be at an impasse, the strong connection between both events through linear oceanic wave propagations originating from the western equatorial Atlantic has the potential to offer predictability of up to 2 months (*42*). Moreover, their relationship with the South Atlantic anticyclone might provide extended predictability if the latter is tied to El Niño Southern Oscillation, for instance. Furthermore, recent advances in deep learning offer a promising avenue for improved prediction of such events. Deep learning techniques (*47*), particularly the application of convolutional neural networks (CNNs), have demonstrated remarkable capabilities in weather forecasting by capturing complex patterns and deciphering intricate relationships within data. For example, using a CNN model, Ham *et al.* (*48*) have shown excellent improvement in the seasonal prediction of El Niño events. Here, we investigate whether a deep learning algorithm, such as a CNN, can predict Atlantic and Benguela Niño/Niña events.

[1]Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway. [2]Climate Simulation and Prediction Division, Euro-Mediterranean Center on Climate Change, Bologna, Italy. [3]Nansen Environmental and Remote Sensing Center, Bergen, Norway. [4]Department of Engineering, University of Cambridge, Cambridge, UK. [5]Laboratoire d'Études en Géophysique et Océanographie Spatiale (LEGOS), CNRS/IRD/UT3/CNES, Toulouse, France. [6]Department of Oceanography, University of Cape Town, Cape Town, South Africa.[7]Nansen-Tutu Centre for Marine Environmental Research, Department of Oceanography, University of Cape Town, South Africa
*Corresponding author. Email: bachelery.marielou@gmail.com

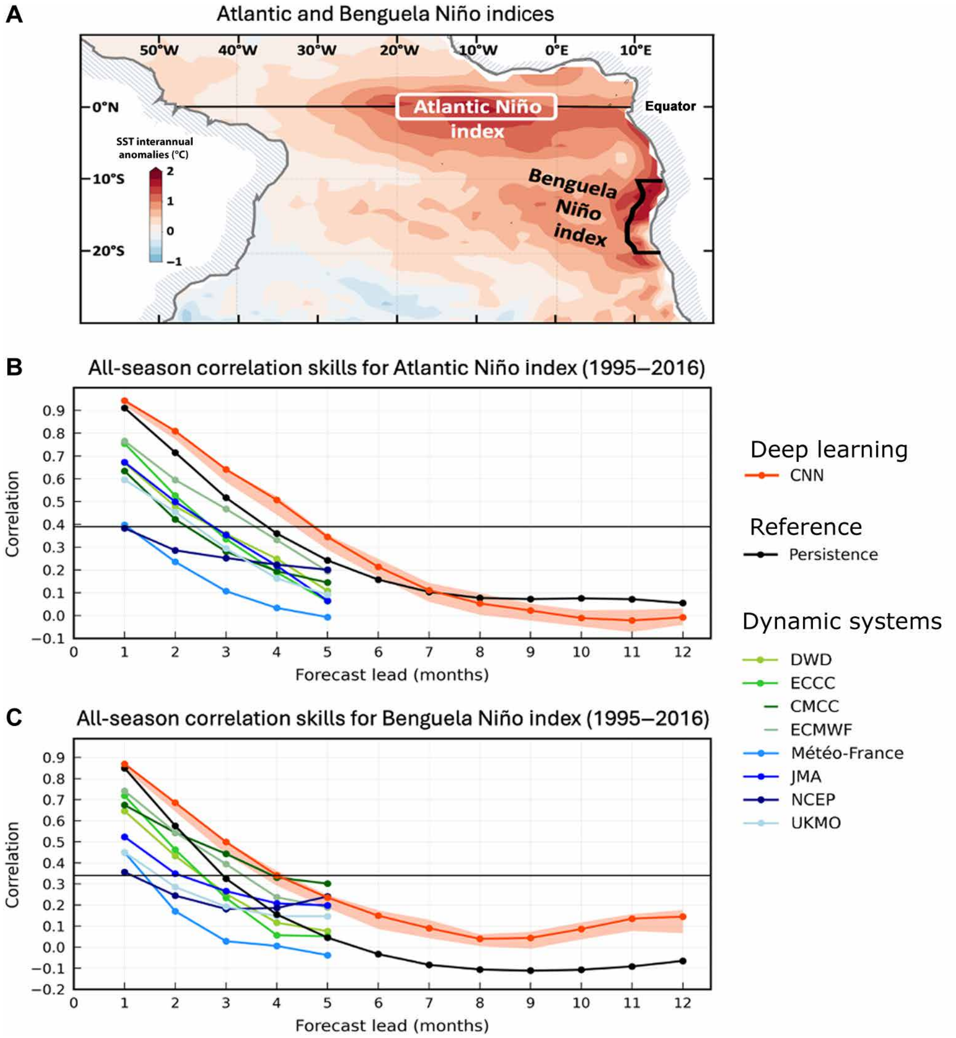

Fig. 1. Prediction performances of state-of-the-art C3S dynamic systems and our deep learning model. (**A**) Map of ORAS5 interannual SST anomalies (°C) during the 2021 Atlantic and Benguela Niño events (May-August 2021). The ANi (20°W-0°E/3°S-3°N) and BNi (10°S-20°S/2°-wide coastal band) domains are outlined in white and black, respectively. (**B**) ANi correlation as a function of lead time (months) for the CNN deep learning model (orange) and the C3S prediction models (blue and green colors). The orange shading shows the highest and lowest correlations estimated over 20 ensemble members of the CNN model. Persistence is shown in black, and the statistical significance threshold [>99%, estimated using a red noise significance method (*83*); see Methods] is indicated by the black horizontal line. The evaluation period is 1995 to 2016, and all models are compared with ORAS5. (**C**) Same as (B) for the BNi.

RESULTS

Assessing the predictability of Atlantic and Benguela Niño indices

Herein, we develop a CNN model to generate seasonal predictions of two indices of Atlantic and Benguela Niño variability (Fig. 1A): the Atlantic Niño index (ANi) and Benguela Niño index (BNi) defined as interannual sea surface temperature (SST) anomalies averaged in the eastern equatorial sector [ATL3 (*49*): 3°N-3°S/20°W-0°E] and in the coastal Angola/Benguela [CABA (*28*): 10°S-20°S/2°-wide coastal band]. The architecture of the CNN statistical model resembles the one described by Ham *et al.* (*48*). While other deep learning architectures have also proved to be skillful, especially for weather forecasts (*50*), here, we use a simple, tailored scheme. This approach will help us explore the potential predictability of tropical Atlantic interannual variability and its extreme events, offering another perspective compared to traditional dynamic model predictions, even with a relatively simple implementation. The CNN model generates monthly forecasts for each index, targeting each month of the year with a lead time ranging from 1 to 12 months. As previously highlighted, interannual SST anomalies in the eastern equatorial and southwest African coastal regions are closely linked to long-wave propagation. Equatorial Kelvin waves and subsequent coastal trapped waves induce substantial changes in vertical and horizontal currents, leading to temperature anomalies within the first 100 m and sea level fluctuations in the eastern equatorial basin (*27*, *28*) and off the African coast (*36*, *37*, *41*). Therefore, the CNN model uses as predictors (input data) maps of monthly interannual surface (SST) and vertically integrated (top 100 m) temperature anomalies over the tropical Atlantic region (15°N-30°S/50°W-20°E). Subseasonal variability has been removed from the predictors (see Methods) to reduce the high-frequency variability and focus the CNN model on the interannual variability, which is the most relevant for predicting the events. Furthermore, we use sequences of maps of both variables during three consecutive months before the forecast start date (see Methods) to depict propagating features and capture the evolution of the interannual events. We train the model using 90 years of reanalysis data from 1900 to 1991. The performances of the CNN model are tested against the Ocean Reanalysis System 5 [ORAS5 (*51*)] dataset covering 1995 to 2016. This period aligns with available seasonal predictions from dynamic forecasting systems used for comparison with the CNN seasonal prediction skills.

The CNN model reveals great success in predicting ANi and BNi variability in terms of both correlation (Fig. 1, B and C) and root mean square error (RMSE; fig. S1 in the Supplementary Materials). It demonstrates predictability extending up to 4-month lead for ANi and 3-month lead for BNi, with correlations surpassing both the persistence (0.39) and statistical significance (0.34). At lead 1, remarkably high correlations of 0.93 (ANi) and 0.88 (BNi) are achieved, gradually decreasing to 0.5 at lead 4 (ANi) and lead 3 (BNi). Compared with the BNi variability, which is influenced by competing local and remote equatorial processes, the ANi variability is more tied to the fully coupled system associated with the Bjerknes feedback. This distinction may account for the higher skills and 1-month-longer forecast horizon observed for ANi compared to BNi.

The CNN model forecasting skills clearly surpass those of the eight state-of-the-art seasonal prediction systems used operationally for the Copernicus Climate Change Service [C3S (*52*)]. Note that in this comparison, the forecast lead time follows the convention of Ham *et al.* (*48*) and Ling *et al.* (*53*) (see Methods). The CNN model demonstrates a superior forecast skill in terms of correlation (Fig. 1, B and C) and RMSE (fig. S1 in the Supplementary Materials) for both indices. Although the relatively recent C3S prediction systems exhibit a substantial improvement in predicting the ANi compared to the existing literature (*12*, *14*, *15*, *20*), none show correlation skills exceeding persistence within a 4-month lead for the Atlantic Niño. The models are even less skillful in predicting the Benguela variability.

To account for which month contributes the most to the all-season correlation skills (Fig. 1), we decomposed correlations into their monthly normalized covariance contributions (see Methods). The CNN model predicts the variability of the Atlantic and Benguela Niño/Niña indices particularly well during their respective peak seasons (Fig. 2), indicating skills in forecasting the onset, evolution, and demise of both events. In the equatorial basin, our monthly normalized covariance decomposition of the CNN model reveals higher performances (above persistence and statistically significant monthly correlations) at 5-month lead time during the June-July period (Fig. 2B), which is associated with the highest ANi variability (Fig. 2A). Good monthly correlation skills are also found from November to March. This season is characterized by a secondary peak of variability (Fig. 2A), most likely associated with the occurrence of Atlantic Niño II events (*54*). In the Angola/Benguela sector, the CNN model (Fig. 2E) shows significant monthly normalized covariances above persistence at leads 1 and 2 only from March/April to July. As persistence drops beyond lead 2 (Fig. 1C), the predictability extends to the months from January to July, the period characterized by strong BNi variability (Fig. 2D) and increased occurrence of Benguela Niño events. Notably, in March-April, the monthly normalized covariances are the highest, with significant values achieved up to 11-month lead. The skills remain low in August-September-October. This season shows the lowest BNi variability (Fig. 2D) with very few extreme events, which could explain why the CNN model did not learn properly. Furthermore, this season is most likely dominated by stochastic atmospheric forcing (wind and heat fluxes), making it impossible for the CNN model to predict. These two aspects suggest that the summer season is characterized by a low signal-to-noise ratio, which reduces predictability as suggested by Li *et al.* (*15*). In addition, a separate monthly normalized covariance decomposition for SST and 100-m-averaged temperature (fig. S2 in the Supplementary Materials) clarifies their respective contributions to the predictive skill. As expected, SST is the most influential, in particular at shorter leads (1 to 2 months). For Benguela events, the 100-m-averaged temperature steadily gains importance at longer leads, whereas for Atlantic events, SST remains consistently dominant, supplemented by notable temperature contributions in specific seasons of the year and notably around the peak in June-July.

It is worth noting that the CNN model shows higher covariances for almost all months compared to the C3S seasonal prediction systems. The improvement is particularly evident for the ANi (Fig. 2C and fig. S3 in the Supplementary Materials). For the BNi, the C3S models have their best significant covariance skills from January to April, with some of the models outperforming the CNN model from lead 3 but showing lower skills for the rest of the months (Fig. 2F and fig. S6 in the Supplementary Materials). The accuracy of the amplitude of the variability of both indices is also substantially improved in the CNN model compared to the dynamic system, as illustrated by the monthly decomposition of the RMSE (figs. S3, S6, and S7 in the Supplementary Materials). Results show that within the first 3-month lead, the seasonal RMSE remains below 1 for all the months for the CNN model, which is largely superior to the skills of any dynamic system.

Beyond the fact that the CNN outperforms predictions made by dynamic systems, the most notable result is that the interannual variability in the eastern equatorial Atlantic and off the Angola/Namibia coast can be forecasted with sufficient lead time to be useful for management services, demonstrating the predictability of these events using a relatively simple statistical approach.

Benguela Niño and Niña events

While the CNN model shows good skills in predicting interannual variability, the true significance lies in its capability to forecast events. We demonstrate this by considering Benguela Niño/Niña events, which are categorized when the BNi exceeds 1 standard deviation (SD) for two consecutive months (Fig. 3A) (*39*, *40*). The modulation of upwelling during Benguela Niños/Niñas not only affects surface temperature but also controls primary production, with negative/positive cascading effects on primary production (*7*) and the reproduction of pelagic fish. A deeper evaluation of the performances of the CNN model underscores its ability in forecasting interannual events compared to normal years.

Overall, both CNN and dynamic systems can predict the occurrence of events with reasonably high accuracy (Fig. 3B). However, the CNN model shows a greater ability in averting false alarms (fig. S8 in the Supplementary Materials), thereby enhancing its credibility in anticipating the onset of the events and providing valuable insights for decision-makers. The CNN model achieves F_1 scores (a composite metric of recall and precision; see Methods) of 0.72, 0.56, and 0.43 for leads 1, 2, and 3, respectively, and remain above persistence skills. Note that differences are significant if larger than 0.11 (see Methods) at leads 1, 2, and 3. Comparison of the CNN model's F_1 scores with

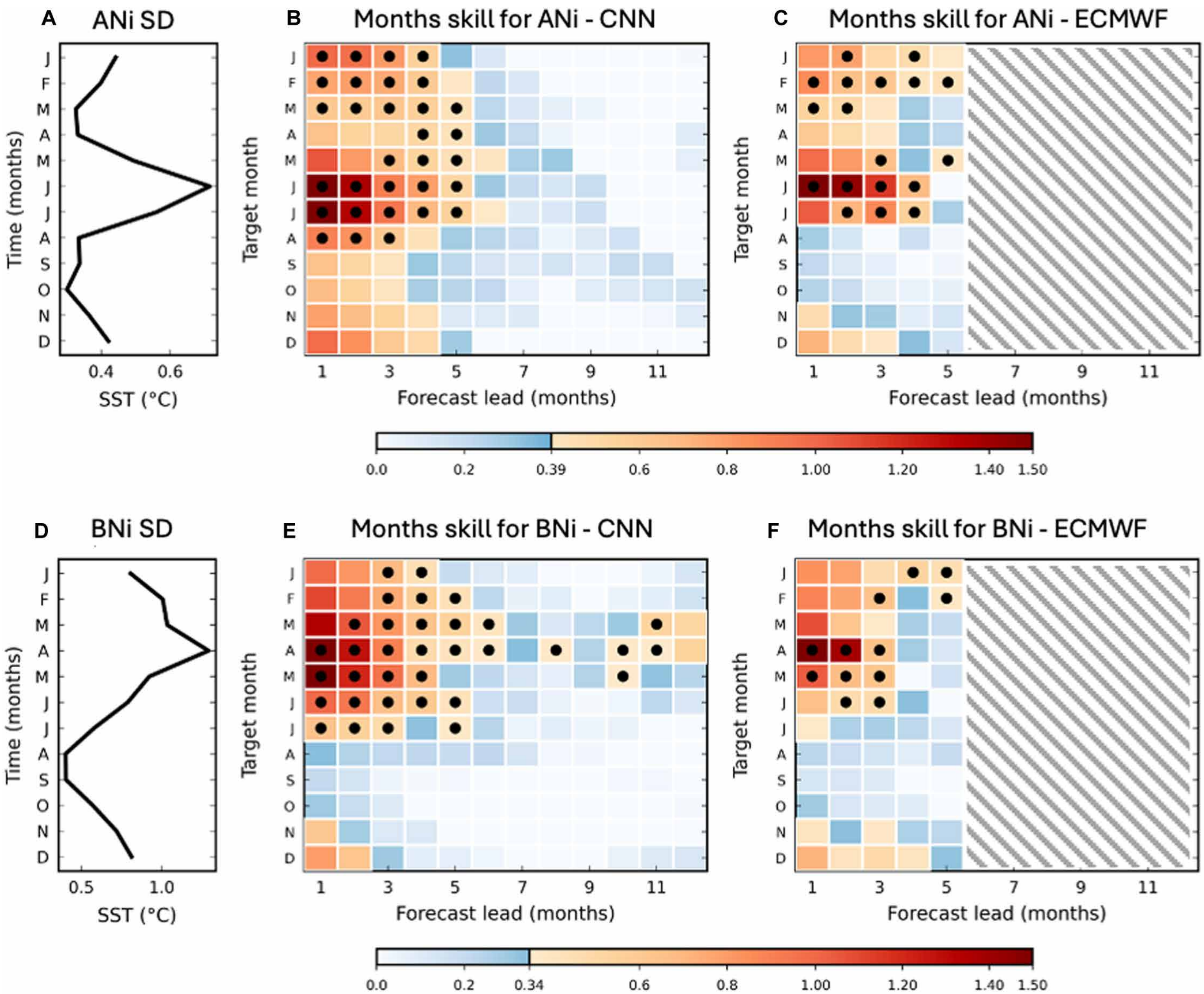

Fig. 2. Monthly prediction performances of the deep learning and ECMWF models. Monthly normalized covariance skills for the Atlantic Niño (top panels; **A** to **C**) and Benguela Niño (bottom panels; **D** to **F**) indices. (Left) Monthly SD of ORAS5 (°C) for the ANi (A) and BNi (D). Normalized covariance decomposition (see Methods) as a function of the lead time (months) for the CNN deep learning model [middle; (B) and (E)] and the ECMWF dynamic model [right; (C) and (F)]. The evaluation period is 1995 to 2016, and all models are compared to ORAS5. Statistically significant values [>99%, estimated using a red noise significance method (*83*); see Methods] are indicated by the red color bar. Monthly significant covariance values above persistence are marked by a black point.

dynamic systems reveals its superiority in predicting events within the first 3-month leads. At lead 3, the recall of CNN is comparable to that of the other dynamic models [such as European Centre for Medium Range Weather Forecasting (ECMWF), CMCC, and UKMO; fig. S8 in the Supplementary Materials]. Meanwhile, dynamic systems predict events more often than observed, resulting in increased false alarms and reduced precision (fig. S8). This discrepancy is also evident in the false positive rate (false positive predictions among total model predictions), with notably lower values for the CNN model for leads 1 to 3, indicating its superior performance. In addition, dynamic systems occasionally predict warm events in cold scenarios and vice versa, a bias not reported for the CNN model predictions.

Notably, the CNN model exhibits superior predictive capabilities for anticipating Benguela Niñas compared to Benguela Niños within the 3-month leads. This contrast is particularly pronounced in lead 3, where the F_1 score for Benguela Niños is only 0.22 compared to 0.65 for Benguela Niñas. This asymmetry suggests potential differences in the phenology of the Benguela Niño and Niña events. It is noteworthy that Benguela Niña subsurface temperature anomalies penetrate the surface layer more easily than those of Benguela Niños, because upwelling waves are associated with a shoaling of the coastal thermocline. This implies better agreement between SST and vertically integrated subsurface temperature and relation to equatorial variability, potentially enhancing predictability.

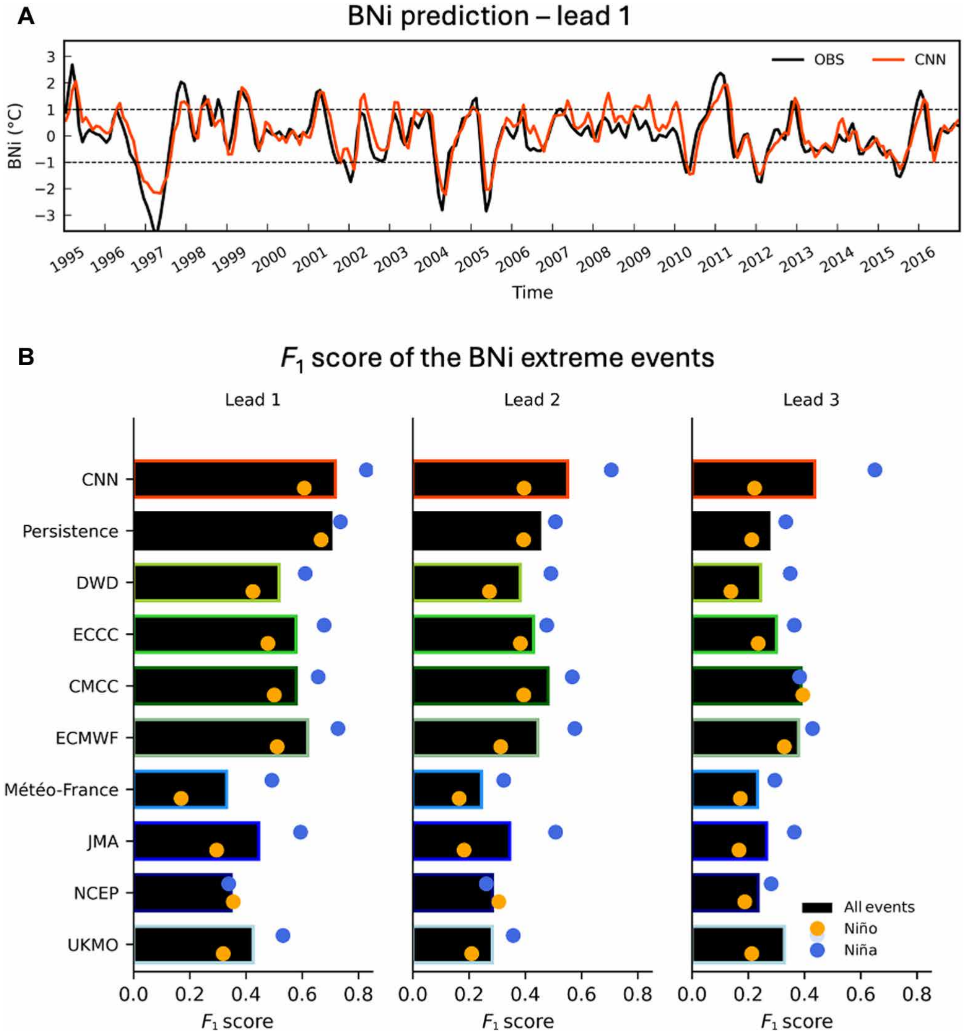

Fig. 3. Skills for predicting the occurrence of Benguela Niños/Niñas. (**A**) BNi time series (°C) for the CNN model predictions (orange) at lead 1. Predictions are compared with ORAS5 (black) over the validation period (1995 to 2016). (**B**) F_1 score (see Methods) for BNi events using the CNN model, the persistence (ORAS5), and the C3S dynamic prediction systems compared with ORAS5. F_1 scores are assessed for leads 1 (left panel) to 3 (right panel) for Benguela Niño (yellow dots), Niña (blue dots), and both events combined (black bars). Contour colors correspond to the color of each model in Fig. 1B. Events are selected when their amplitude exceeds 1 SD for two consecutive months.

Precursors of the Atlantic and Benguela Niño/Niña events

Beyond its success in prediction, deep learning can be a valuable tool for analyzing the complex dynamics and phenology of climate extremes. To investigate hotspots within precursors, especially during the onset of an event, we computed gradient sensitivity (or heatmaps; see Methods) for Atlantic and Benguela events with a lead time of 3 to 1 months (Fig. 4). Heatmaps of the gradients highlight the regions of importance within the predictors when computing the CNN model weights. Higher values highlight the location where predictors make a larger contribution to the predictands. The skill in predicting Atlantic Niño events is connected to westerly wind bursts and increased equatorial heat content that occur 3 to 4 months prior that are redistributed through planetary equatorial wave dynamics to off-equatorial regions (*29*, *55*). We illustrate this by examining the skillful prediction of the 1996 Atlantic Niño event, which peaked in June and was the strongest event over 1995–2016. The heatmaps identify the importance of precursor heat content signals in the eastern and western equatorial Atlantic at 3-month lead, matching both the buildup of equatorial heat content and the impact of westerly wind anomalies in the west (Fig. 4A). There are signatures of coastal trapped waves along the west African coast at 2-month lead (Fig. 4B) and of the reflection of equatorial Kelvin waves into equatorial Rossby waves at 1-month lead (Fig. 4C). These features coincide with the redistribution of equatorial heat content and further warming in the east. A specificity of the 1996 Atlantic Niño is its connection to the cold conditions in the tropical Pacific (*28*, *56*). In late 1995/early 1996, changes in the Atlantic Walker circulation, characterized by anomalous subsidence over South America, led to enhanced westerlies that forced downwelling equatorial Kelvin waves. At 3-month lead (Fig. 4A), the heatmap captures precursors in the western part of the equatorial basin, hugging the South American coasts, which is consistent with this storyline.

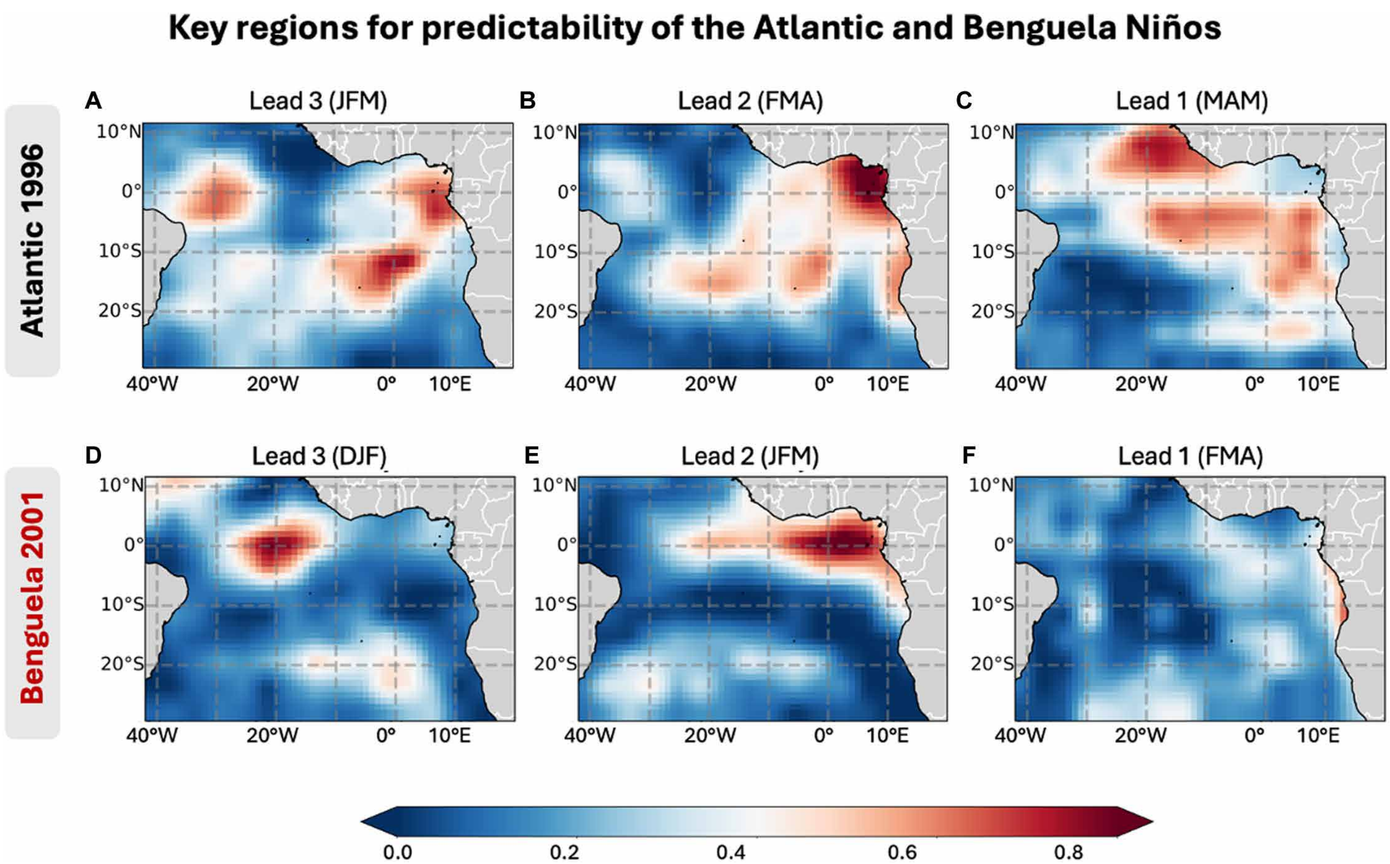

Fig. 4. CNN predictor hotspots for the 1996 Atlantic Niño and 2001 Benguela Niño forecasts. (Top) Heatmaps for the June 1996 Atlantic Niño forecast at (**A**) lead 3 (i.e., using precursors from January-February-March), (**B**) lead 2 (i.e., with precursors from February-March-April), and (**C**) lead 1 (i.e., with precursors from March-April-May). (Bottom; **D** to **F**) Same for the May 2001 Benguela Niño forecast.

Equatorial wave dynamics also served as an important precursor for skillful predictions of Benguela Niño/Niña events, as shown for the 2001 event (Fig. 4, D to F). The CNN nicely predicts this well-documented event, which peaked in May (*38*, *41*, *43*) (Fig. 3A). At the 1-month lead (Fig. 4F), the heatmap reveals the importance of the Angolan and Namibian coast, the region of occurrence of the Benguela Niño/Niña events. One month before (at lead 2; Fig. 4E), the area of importance extends equatorward, north of the Angola-Benguela coastal region, and along the equator, highlighting the equatorial and coastal waveguides. Last, at lead 3 (Fig. 4D), the heatmap captures that precursors originate from temperature anomalies in the west-central equatorial basin, where initial equatorial Kelvin waves are generated.

The predictability of each Benguela or Atlantic event may arise from different regions in the precursors, activated by distinct external forcings that are reflected in the hotspots depicted by the heatmap analysis. For instance, the heatmaps of the 3-month lead forecasts (Fig. 4D) not only highlight the west-central equatorial basin as an important region for the prediction of the coastal warming in May 2001 but also depict the central south tropical basin. Variations in the position and the amplitude of the South Atlantic high-pressure system most likely trigger changes in the wind circulation, evaporation, and consequently, temperature in the central basin. These results therefore underscore the importance of the South Atlantic high-pressure system as a driver of the Benguela events and point to it as a good potential predictor for improving their forecasts. Resolving the deficiencies of dynamic models in capturing the South Atlantic anticyclone (*57*, *58*) might thus play a key role in achieving a more accurate representation of the dynamics of the south tropical Atlantic (*59*). Other examples illustrating the diversity among events include the Benguela Niños documented to be initiated by local wind stress anomalies along the west African coast, potentially associated with changes in the South Atlantic anticyclone (*27*–*30*) or associated with change in salinity near the Congo river (*60*, *61*). Remarkably, this peculiar dynamic was also well captured when analyzing the physical precursor from CNN model predictions, with 3-month predictions having their hotspots of predictability located in the central basin, extending to the tip of the African continent or near the Congo mouth (fig. S9 in the Supplementary Materials).

Forecasting the 2021 Atlantic and Benguela Niño events

In the context of this study, the effective development of predictions for Atlantic and Benguela Niños and Niñas represents an opportunity to establish a warning system for vulnerable communities. To achieve this, it is essential to automate the forecasting process using the latest available data and to make the forecasts easily accessible on the internet. This section outlines an attempt to realize this vision, with the goal of providing real-time, freely available forecasts to empower communities in anticipating and preparing for these extreme climatic events. The 2021 Atlantic and Benguela Niño events, owing to their substantial magnitude, impact, and interest in the scientific community (*15*, *29*, *62*), are used here as a case study. The 2021 Atlantic Niño event, which peaked in July, was the warmest equatorial event since 1982 (*29*), substantially affecting rainfall over African countries and India (*63*). Concurrently, the intensity and late onset of the May 2021 Benguela Niño yielded the weakest boreal summer primary production off Angola since 2002 (*64*).

To develop predictions of recent events (including the 2021 events), we use the CNN model here with predictors from the observation-based ARMOR3D dataset. Compared to reanalyses used to train and evaluate the CNN model, ARMOR3D is a dynamically updated dataset, incorporating the latest near-real-t ime altimetry maps. This allows for operational utilization in developing a pioneering web-based warning system for both the eastern equatorial and southwestern African regions. It also allows the introduction of an additional and entirely new source of data to test the forecast performances of our deep learning model. The CNN model accurately predicts the equatorial event 4 months in advance, and the amplitude of the ANi forecasts effectively matches well with the observations (Fig. 5A). Despite the CNN model's lower success in predicting the June/July BNi, it accurately predicts the Benguela warming 1 and 4 months in advance (Fig. 5B). The timing of the interannual BNi forecasts closely resemble that of the satellite observations, although there is an underestimation of the amplitude of the June/July events in both 2020 and 2021. Li *et al.* (*15*) showed that the state-of-the-art dynamic prediction systems did not capture the occurrence of the 2021 Atlantic and Benguela Niño events. In contrast, our accurate tropical Atlantic forecasts using ARMOR3D data instill confidence in the CNN model's capabilities, support its potential for operational forecasting, and offer promising avenues for the moni-toring of critical marine ecosystems.

DISCUSSION

Decades of intensive research have been dedicated to enhancing predictions for the tropical Atlantic. However, as for now, the Atlantic and Benguela Niños remain poorly predicted using dynamic seasonal prediction systems. This study presents the potential of deep learning in the prediction of the tropical Atlantic interannual variability and the Atlantic and Benguela Niño/Niña events. Results show that the CNN model developed herein provides reliable forecasts for both events up to 3 months ahead, surpassing the skills of state-of-the-art dynamic forecasting systems. The progress achieved using deep learning is higher for the Atlantic index than for the Benguela index.

While our CNN model demonstrates good predictive capabilities, its interpretability remains a challenge. By their very nature, deep learning models make it difficult to fully understand the underlying mechanisms driving the predictions. However, the increased prediction skill of the CNN model might arise from its ability to learn complex spatial and temporal patterns in the data that serve as precursors to Atlantic and Benguela Niño events. By learning directly from preprocessed observational data, the CNN model can overcome some of the recurring errors in global climate models, such as biases and erroneous variability. Also, by not relying only on explicitly modeled physical processes, the CNN model can bypass systemic issues related to parameterizations and resolution constraints, potentially capturing subtle signals and nonlinear relationships that are not well represented in dynamic models.

Analyzing the gradient sensitivity of the CNN model enables one to address the challenge of interpretability and identify the potential precursors and intricate mechanisms influencing the tropical Atlantic predictability across different lead times. The predictability of both

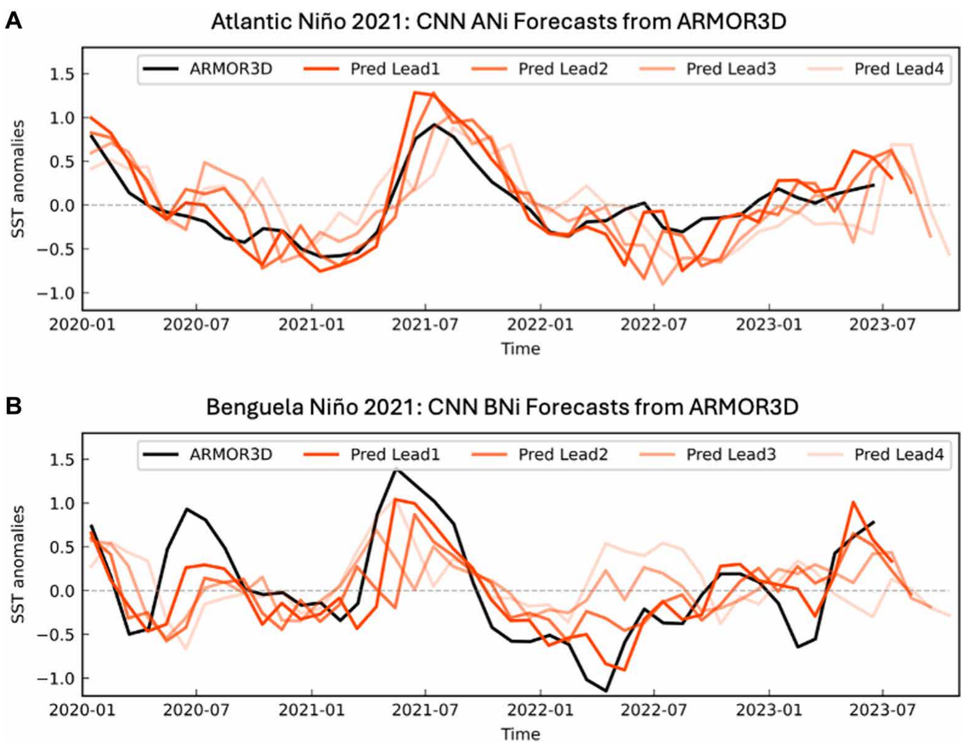

Fig. 5. 2021 Atlantic and Benguela Niño forecasts. (**A**) Forecasts of the ANi (°C) using the CNN model from lead 1 (deep orange) to lead 4 (light orange). Forecasts are estimated from and compared to the Copernicus ARMOR3D datasets. (**B**) Same as (A) but for the BNi.

events arises from well-documented mechanisms, predominantly associated with long-wave dynamics. Temperature signals associated with changes in the South Atlantic anticyclone are also captured as a notable precursor to the Benguela events, potentially extending their prediction horizon further. Identifying these precursors is essential first for validating the CNN model's credibility. It can help understand why dynamic seasonal prediction systems are deficient, providing clues on mechanisms not well represented within them.

While this study represents a substantial step forward in the application of deep learning for climate prediction, there is an ample room for improvement. First, the relatively short validation period (1995 to 2016) may limit the assessment of the model's robustness, and when possible, future studies should evaluate the CNN model over longer periods to confirm and enhance its predictive capabilities as the model's skill may vary when applied to different periods or under different climatic conditions. Nevertheless, the lightweight and computationally efficient design of the CNN model allows for easy retraining with the most recent data. This capability enables the model to continually update and potentially improve its predictive accuracy, thereby remaining relevant for operational use. Second, because the objective was to assess the potential of predictability of the Atlantic and Benguela events rather than to design a distinct deep learning architecture, the CNN model developed here remains relatively simple and can be further refined and enhanced. A notable limitation is the scarcity of data available for training the model, which often constrains its predictive capabilities. Very popular, transfer learning techniques can augment data volume, potentially enhancing the CNN model's performance. We explored transfer learning by initially training the model on CMIP6 hindcasts before fine-tuning it with reanalysis data. However, the CMIP6 models exhibit poor qualities in simulating the tropical Atlantic interannual variability and, particularly, the characteristics of the Benguela events, thereby degrading the forecast accuracy. Although transfer learning could not be applied here, the rapid evolution of machine learning techniques should continue to offer promising avenues for improving the model's skills. Future work may also involve exploring more complex neural network architectures, such as ConvLSTM (*65*, *66*), multimodal approaches (*67*), and transformers and attention mechanisms (*68*–*70*), optimizing hyperparameters, or integrating an extreme-focused loss function, which applies extra penalties to extreme events (*71*–*74*), to improve predictive accuracy and further enhance the model's ability to forecast the Atlantic and Benguela Niño events. In addition, incorporating additional climate data sources, variables (wind patterns or sea level pressure), or external drivers (climate indices such as ENSO and/or the Indian Ocean Dipole) as compact representations of remote influences could substantially improve the model's ability to capture the phenology of the events and extend prediction horizons.

METHODS

Design of the CNN model

A CNN is designed for image analysis (*47*). It consists of multiple convolutional, pooling, and fully connected layers. Herein, the CNN statistical model's architecture resembles the one described by Ham *et al.* (*48*). It is built upon three convolutional layers interspersed by two max pooling layers. The last convolutional layer is connected to neurons in the fully connected layer, which, in turn, is linked to the final output. Both the total numbers of convolutional filters and neurons in the fully connected layer are set to 30. During the training process, a minibatch size of 20 is used for each epoch and the model undergoes 125 epochs. We define learning as a regression problem, with the dimensionality of the input set to a height of 64, a width of 44, and six channels for the input images. The optimizer used is RMSProp with default hyperparameters (betas = 0.9 and 0.999; epsilon = 1×10^{-7}). The learning rate remains fixed to 5×10^{-5}. A dropout rate of 0.25 is applied to prevent overfitting. Note that apart from the optimizer, the hyperparameter settings were adapted from the architecture by Ham *et al.* (*48*) as they demonstrated an improved predictive skill. A targeted hyperparameter tuning study, including a comparison of optimizers (RMSProp, Adam, and Nadam), was conducted to further ensure that the chosen settings are among the best possible combinations for our study (figs. S10 and S11). The full details of this hyperparameter tuning analysis are provided in the Supplementary Materials.

Central to the CNN framework, the convolutional process involves the systematic application of small learnable filters to input data. These filters slide over the data, performing element-wise multiplications and producing a feature map that highlights relevant patterns. This process captures hierarchical features, enabling the network to progressively acquire informative representations. The resulting feature maps are subject to nonlinear activation functions and optional pooling operations, contributing to the network's ability to learn and generalize from input data. As by Ham *et al.* (*48*), a hyperbolic tangent activation function is used in the CNN model. Filter dimensions are defined as 3 by 3 for all the convolutional layers. Padding is used to fill vacant spaces with zeros, preserving the feature maps' dimensions intact.

Forecasts of the Benguela Niño/Niña index (BNi) and Atlantic Niño/Niña index (ANi)

The CNN model is designed to forecast the ANi and BNi (predictands). The ANi corresponds to the interannual SST anomalies averaged over the eastern equatorial Atlantic [ATL3 box (*49*): 20°W-0°E/3°S-3°N]. The BNi is defined as the interannual SST anomalies averaged along the coast of West Africa in the Coastal Angola-Benguela area [CABA box (*28*): 20°S-10°S;2°-wide coastal band]. The CNN model uses monthly-averaged surface and 100-m-integrated temperature anomaly maps of the tropical Atlantic region (15°N-30°S/50°W-20°E) for three consecutive months as predictors. These data capture the propagation of long waves along the equator and down the west coast of Africa. While the sea level is traditionally used to assess long-wave dynamics, it provides little information associated with higher, slower baroclinic modes, thus offering less range (memory) to the predictors. Note that in the early stages of the model development, other variables were considered as predictors (including sea level and wind stress anomalies). While incorporating wind stress anomalies as a predictor could have helped the CNN model account for the ocean-atmosphere interactions and coupling between the equatorial and coastal events, it did not improve the prediction skill. Possible reasons include the limited training data and noise levels in the wind dataset, which may have hindered the model's ability to learn effectively.

The CNN model generates forecasts for each "target" month of the year. Each forecast uses the latest monthly available data (predictors). The "lead time" represents the number of months between the latest observed data and the target month for the forecast. For instance, to produce a forecast for January with a lead time of 1 month, the CNN model uses data from the three preceding months (October, November, and December) of the previous year. For a lead time of 2 and a forecast for January, the model uses data from September, October, and November of the previous year and so on for subsequent lead times. The forecast target periods for all lead times and target months range from January 1995 to December 2016. This evaluation period is determined by the availability of the seasonal dynamic forecasting system outputs used for comparison (see the "Reanalyses, observed datasets, and seasonal dynamic forecasts" section for the description of the dynamic forecast systems). To improve the reliability of the results, we do an ensemble learning, creating 20 variations of the CNN model. Each of these 20 models is trained for every target month and lead time, resulting in a set of 20 distinct models for each scenario. The final forecasts are generated by taking the ensemble mean of these multiple model prediction estimates.

CNN heatmaps

Heatmaps generated using the Grad-CAM (gradient-weighted class activation mapping) (*75*) technique are used to assess the decision-making process of the CNN. Gradients are extracted from the third convolutional layer of the trained model. These heatmaps provide a visual aid to understand the features and spatial regions that the CNN model relies on for accurate predictions, thereby informing on the physical precursors influencing the variability of the ANi and BNi. Higher gradients indicate a stronger influence of the corresponding features on the variability observed in the indices.

Reanalyses, observed datasets, and seasonal dynamic forecasts

The CNN model is trained using maps of surface and subsurface temperature anomalies from the CMCC Historical Ocean Reanalysis [CHORE_RL (*76*)] (Table 1). The reanalysis provides monthly-averaged outputs over the period of 1900 to 2010 at a resolution of 1/2° by 1/2° and uses data from the ECMWF ERA-20C historical atmospheric reanalysis as surface forcing (*77*). It assimilates vertical profile data and is nudged to monthly Hadley Centre global sea ice and SST reconstructed fields (*78*). Only CHORE_RL outputs from 1900 to 1991 have been used for the training of the CNN model. This was done to ensure a sufficiently long validation period independent of the training period.

To evaluate forecast performance, we use monthly SST and subsurface temperature data from 1993 to the present from ORAS5 (*51*) (Table 1). The reanalysis combines model data with observations and provides outputs at a resolution of 1/4° by 1/4°. From 1993 to 2014, outputs are extracted from the consolidated product, while from 2015 onward, the operational product is used. The consolidated and operational ORAS5 data differ in their atmospheric forcing and the ocean observations used to generate the two products.

Seasonal predictions from the eight systems of C3S (*52*) (Table 1) are used to benchmark the prediction skills of the CNN model over 1995 to 2016 (Table 2). These state-of-the-art seasonal prediction systems provide monthly SST forecasts with lead times from 0 up to 5 months at a spatial resolution of 1° by 1°. Note that the lead times have been defined as by Ham *et al.* (*48*) and Ling *et al.* (*53*).

Forecasts for the 2021 Atlantic and Benguela Niño events (Fig. 5) are generated using predictor temperature maps from the ARMOR3D product (*79*, *80*) (Table 1). ARMOR3D is a 1/4° by 1/4° observation-based product that integrates satellite data (sea level anomalies and SST) with in situ vertical profiles of temperature and salinity from Argo, CTD, XBTs, etc., using statistical methods. We used the near-real-time ARMOR3D L4 analysis, which is based on near-real-time altimetry maps (*81*) from 2019 to 2024. Note that the climatology was calculated from the multiyear reprocessed ARMOR3D with data available from 1993 to 2022.

Table 1. Overview of the datasets. Name, types, periods, lead time, and goal of the datasets.

Name	Type	Period	Lead time	Goal
CHORE_RL	Reanalysis	1900–1991	x	Training set
ORAS5	Reanalysis	1993–2022	x	Test set
C3S	Seasonal forecasts	1995–2016	6 months	Validation
ARMOR3D	Observation-based product	2019–2023	x	Prediction exercise

Table 2. Details of C3S prediction systems. Information (institution, name, system, and ensemble size) on the C3S seasonal prediction systems used in this study.

Institution	Name	System	Ensemble size
Deutscher Wetterdienst	DWD	21	30
Environment and Climate Change Canada	ECCC	3	10
Centro Euro-Mediterraneo sui Cambiamenti Climatici	CMCC	35	40
European Centre for Medium-Range Weather Forecast	ECMWF	seas5	51
Météo-France	Météo-France	8	25
Japan Meteorological Agency	JMA	3	10
National Centers for Environmental Prediction	NCEP	2	28
UK Met Office	UKMO	600	28

All datasets used for training, validation, and comparison with the dynamic forecast systems are first linearly interpolated over the same grid at a degree resolution of 1° by 1°. A common land mask is also used for all datasets.

Interannual anomalies

Interannual anomalies for all datasets are computed as by Bachèlery *et al.* (*44*). First, monthly anomalies are calculated by estimating monthly means and subtracting the monthly climatological mean state. Note that for the training set, the climatology is calculated over the period of 1900–1991, while for the test set, the seasonal cycle is estimated over the period of 1993 to 2022. The climatologies of the dynamic prediction systems are lead time dependent. Subseasonal variations are filtered out using a weighted average, with a 1-2 decentered running weighted average applied to mitigate contamination by future data as a caution by Liu *et al.* (*82*). This methodological choice ensures that the model focuses on the interannual variability, which is most relevant for predicting Atlantic and Benguela Niño events. Sensitivity tests show that filtering out higher-frequency variability considerably improves the CNN model ability to learn the interannual patterns associated with the events. Last, data are not detrended.

Forecast performance metrics

The forecasting skills of the CNN and dynamic system models are first evaluated for each lead time (l) against ORAS5 data by examining the correlation coefficients (C_l) and the normalized RMSE (RMSE_l) over the entire validation period (Fig. 1 and figs. S1 and S3 in the Supplementary Materials).

$$C_l = \frac{\sum_{i=1}^{N} O_i \cdot P_{i,l}}{\sqrt{\left(\sum_{i=1}^{N} O_i^2\right) \cdot \left(\sum_{i=1}^{N} P_{i,l}^2\right)}} \quad (1)$$

$$\text{RMSE}_l = \sqrt{\frac{\sum_{i=1}^{N} (O_i - P_i)^2}{\sum_{i=1}^{N} O_i^2}} \quad (2)$$

where O and P are the observed (ORAS5) and predicted ANi or BNi values, respectively, and l and N are the lead time and the length of the time series ($N = 22 \times 12$), respectively.

Then, to evaluate the performances for the different months of the year (Fig. 2 and figs. S4 to S7 in the Supplementary Materials), the all-month correlation and RMSE metrics are broken down into each calendar month contribution (M). For the correlation, this involves normalizing the monthly covariances by the SD of the all-month time series such that the sum of these monthly contributions corresponds exactly to the all-month correlation (presented in Fig. 1) $C_l = \frac{1}{12}\sum_{M=1}^{12} C_{l,M}$ and $\text{RMSE}_l = \frac{1}{12}\sum_{M=1}^{12} \text{RMSE}_{l,M}$, with

$$C_{l,M} = \frac{\left(\sum_{y=1995}^{2016} O_{y,M} \cdot P_{l,y,M}\right)/n}{\sqrt{\left(\sum_{y=1995}^{2016} \sum_{m=1}^{12} O_{y,m}^{\ 2}\right)/N \cdot \left(\sum_{y=1995}^{2016} \sum_{m=1}^{12} P_{l,y,m}^{\ 2}\right)/N}} \quad (3)$$

$$\text{RMSE}_{l,\text{M}} = \sqrt{\frac{\left[\sum_{y=1995}^{2016} \left(O_{y,\text{M}} - P_{l,y,\text{M}}\right)^2\right]/n}{\left(\sum_{y=1995}^{2016} \sum_{m=1}^{12} O_{y,m}^{\ 2}\right)/N}} \quad (4)$$

where y, m/M, and n are the year, the month, and the number of years ($n = 22$), respectively.

The statistically significant correlation threshold at the 99% confidence level is determined from the probability distribution functions of 10,000 instances of the correlation coefficient between two red noise samples with the same autoregressive characteristics as the original signals (*83*).

To measure the success of the models in predicting Benguela Niño/Niña events, we estimate F_1 scores, recall, precision, and false positive rates. The F_1 score is widely used in the imbalanced classification problem. It is the harmonic mean of the precision and recall. It is defined by

$$F_1 \text{ score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (5)$$

with precision and recall defined by

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

Here, TP (true positive), TN (true negative), FP (false positive), and FN (false negative) represent correct and incorrect detection of Benguela Niño/Niña events.

In addition, we calculate the false positive rate, which quantifies the probability of raising a false alarm. It is defined as

$$\text{False positive rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (8)$$

Note that events are categorized when the BNi exceeds 1 SD for two consecutive months. The significance of the F_1 score differences between the CNN model and the average of the seasonal prediction systems is estimated using a bootstrap method based on the generation of 10,000 shuffled samples (*84*).

Supplementary Materials

This PDF file includes:
Supplementary Notes S1 to S5
Figs. S1 to S11

REFERENCES AND NOTES

1. M. Rouault, P. Florenchie, N. Fauchereau, C. J. C. Reason, South East tropical Atlantic warm events and southern African rainfall. *Geophys. Res. Lett.* **30**, 8009 (2003).
2. P. Brandt, G. Caniaux, B. Bourlès, A. Lazar, M. Dengler, A. Funk, V. Hormann, H. Giordani, F. Marin, Equatorial upper-ocean dynamics and their interaction with the West African monsoon. *Atmos. Sci. Lett.* **12**, 24–30 (2011).
3. K. Lutz, J. Jacobeit, J. Rathmann, Atlantic warm and cold water events and impact on African west coast precipitation. *Int. J. Climatol.* **35**, 128–141 (2015).
4. M. Rouault, Bi-annual intrusion of tropical water in the northern Benguela upwelling. *Geophys. Res. Lett.* **39**, L12606 (2012).
5. P. M. S. Monteiro, A. van der Plas, V. Mohrholz, E. Mabille, A. Pascall, W. Joubert, Variability of natural hypoxia and methane in a coastal upwelling system: Oceanic physics or shelf biology? *Geophys. Res. Lett.* **33**, L16614 (2006).

6. P. M. S. Monteiro, A. K. van der Plas, J.-L. Mélice, P. Florenchie, Interannual hypoxia variability in a coastal upwelling system: Ocean–shelf exchange, climate and ecosystem-state implications. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **55**, 435–450 (2008).
7. M.-L. Bachèlery, S. Illig, I. Dadou, Forcings of nutrient, oxygen, and primary production interannual variability in the southeast Atlantic Ocean. *Geophys. Res. Lett.* **43**, 8617–8625 (2016).
8. C. Wang, An overlooked feature of tropical climate: Inter-Pacific-Atlantic variability. *Geophys. Res. Lett.* **33**, L12702 (2006).
9. H. Ding, N. S. Keenlyside, M. Latif, Impact of the equatorial Atlantic on the El Niño Southern Oscillation. *Climate Dynam.* **38**, 1965–1972 (2012).
10. W. Johns, S. Speich, M. Araujo, M. Balmaseda, P. Chang, P. Dandin, K. Hill, N. Keenlyside, J. Knight, Y. Kushnir, M. McPhaden, I. Richter, C. Robinson, R. R. Rodrigues, J. O. Schmidt, A. Simmons, N. Smith, S. Stripling, T. Tanhua, M. Visbeck, Tropical Atlantic Observing System (TAOS) review report (CLIVAR-01/2021, 2021); www.clivar.org/documents/tropical-atlantic-observing-system-taos-review-report.
11. I. Richter, T. Doi, S. K. Behera, N. Keenlyside, On the link between mean state biases and prediction skill in the tropics: An atmospheric perspective. *Climate Dynam.* **50**, 3355–3374 (2018).
12. X. Li, M. H. Bordbar, M. Latif, W. Park, J. Harlaß, Monthly to seasonal prediction of tropical Atlantic sea surface temperature with statistical models constructed from observations and data from the Kiel Climate Model. *Climate Dynam.* **54**, 1829–1850 (2020).
13. P. Oettli, C. Yuan, I. Richter, "10 - The other coastal Niño/Niña—The Benguela, California, and Dakar Niños/Niñas," in *Tropical and Extratropical Air-Sea Interactions*, S. K. Behera, Ed. (Elsevier, 2021), pp. 237–266.
14. R. Wang, L. Chen, T. Li, J.-J. Luo, Atlantic Niño/Niña prediction skills in NMME models. *Atmosphere* **12**, 803 (2021).
15. X. Li, W. Tan, Z.-Z. Hu, N. C. Johnson, Evolution and prediction of two extremely strong Atlantic Niños in 2019–2021: Impact of Benguela warming. *Geophys. Res. Lett.* **50**, e2023GL104215 (2023).
16. N. Keenlyside, Y. Kosaka, N. Vigaud, A. W. Robertson, Y. Wang, D. Dommenget, J.-J. Luo, D. Matei, "Basin interactions and predictability," in *Interacting Climates of Ocean Basins: Observations, Mechanisms, Predictability, and Impacts* (Cambridge Univ. Press, 2020), pp. 258–292.
17. T. N. Stockdale, D. L. T. Anderson, M. A. Balmaseda, F. Doblas-Reyes, L. Ferranti, K. Mogensen, T. N. Palmer, F. Molteni, F. Vitart, ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Clim. Dyn.* **37**, 455–471 (2011).
18. I. Richter, S. K. Behera, T. Doi, B. Taguchi, Y. Masumoto, S.-P. Xie, What controls equatorial Atlantic winds in boreal spring? *Climate Dynam.* **43**, 3091–3104 (2014).
19. T. Dippe, R. J. Greatbatch, H. Ding, On the relationship between Atlantic Niño variability and ocean dynamics. *Climate Dynam.* **51**, 597–612 (2018).
20. F. Counillon, N. Keenlyside, T. Toniazzo, S. Koseki, T. Demissie, I. Bethke, Y. Wang, Relating model bias and prediction skill in the equatorial Atlantic. *Climate Dynam.* **56**, 2617–2630 (2021).
21. Y. Chen, X. Yan, Y. Tang, Q. Song, Z. Shen, Y. Wu, Decline in Atlantic Niño prediction skill in the North American multi-model ensemble. *Commun. Earth Environ.* **5**, 524 (2024).
22. B. Huang, Z.-Z. Hu, B. Jha, Evolution of model systematic errors in the Tropical Atlantic Basin from coupled climate hindcasts. *Climate Dynam.* **28**, 661–682 (2007).
23. Z. Xu, P. Chang, I. Richter, W. Kim, G. Tang, Diagnosing southeast tropical Atlantic SST and ocean circulation biases in the CMIP5 ensemble. *Climate Dynam.* **43**, 3123–3145 (2014).
24. Z. Xu, M. Li, C. M. Patricola, P. Chang, Oceanic origin of southeast tropical Atlantic biases. *Climate Dynam.* **43**, 2915–2930 (2014).
25. I. Richter, S.-P. Xie, S. K. Behera, T. Doi, Y. Masumoto, Equatorial Atlantic variability and its relation to mean state biases in CMIP5. *Climate Dynam.* **42**, 171–188 (2014).
26. Z.-Z. Hu, B. Huang, Physical Processes associated with the tropical Atlantic SST gradient during the anomalous evolution in the Southeastern Ocean. *J. Climate* **20**, 3366–3378 (2007).
27. J. F. Lübbecke, C. W. Böning, N. S. Keenlyside, S.-P. Xie, On the connection between Benguela and equatorial Atlantic Niños and the role of the South Atlantic Anticyclone. *J. Geophys. Res. Oceans.* **115**, doi.org/10.1029/2009JC005964 (2010).
28. S. Illig, M.-L. Bachèlery, J. F. Lübbecke, Why do Benguela Niños lead Atlantic Niños? *J. Geophys. Res. Oceans.* **125**, e2019JC016003 (2020).
29. S. Illig, M.-L. Bachèlery, The 2021 Atlantic Niño and Benguela Niño events: External forcings and air–sea interactions. *Clim. Dynam.* **62**, 679–702 (2024).
30. I. Richter, S. K. Behera, Y. Masumoto, B. Taguchi, N. Komori, T. Yamagata, On the triggering of Benguela Niños: Remote equatorial versus local influences. *Geophys. Res. Lett.* **37**, doi.org/10.1029/2010GL044461 (2010).
31. T. Junker, M. Schmidt, V. Mohrholz, The relation of wind stress curl and meridional transport in the Benguela upwelling system. *J. Mar. Syst.* **143**, 1–6 (2015).
32. S. Illig, B. Dewitte, N. Ayoub, Y. du Penhoat, G. Reverdin, P. De Mey, F. Bonjean, G. S. E. Lagerloef, Interannual long equatorial waves in the tropical Atlantic from a high-resolution ocean general circulation model experiment in 1981–2000. *J. Geophys. Res. Oceans.* **109**, doi.org/10.1029/2003JC001771 (2004).
33. S. Illig, D. Gushchina, B. Dewitte, N. Ayoub, Y. du Penhoat, The 1996 equatorial Atlantic warm event: Origin and mechanisms. *Geophys. Res. Lett.* **33**, doi.org/10.1029/2005GL025632 (2006).
34. G. R. Foltz, M. J. McPhaden, Abrupt equatorial wave-induced cooling of the Atlantic cold tongue in 2009. *Geophys. Res. Lett.* **37**, doi.org/10.1029/2010GL045522 (2010).
35. M. Martín-Rey, A. Lazar, Is the boreal spring tropical Atlantic variability a precursor of the Equatorial Mode? *Climate Dynam.* **53**, 2339–2353 (2019).
36. P. Florenchie, J. R. E. Lutjeharms, C. J. C. Reason, S. Masson, M. Rouault, The source of Benguela Niños in the South Atlantic Ocean. *Geophys. Res. Lett.* **30**, doi.org/10.1029/2003GL017172 (2003).
37. P. Florenchie, C. J. C. Reason, J. R. E. Lutjeharms, M. Rouault, C. Roy, S. Masson, Evolution of interannual warm and cold events in the southeast Atlantic ocean. *J. Climate* **17**, 2318–2334 (2004).
38. M. Rouault, S. Illig, C. Bartholomae, C. J. C. Reason, A. Bentamy, Propagation and origin of warm anomalies in the Angola Benguela upwelling system in 2001. *J. Mar. Syst.* **68**, 473–488 (2007).
39. M. Rouault, S. Illig, J. Lübbecke, R. A. I. Koungue, Origin, development and demise of the 2010–2011 Benguela Niño. *J. Mar. Syst.* **188**, 39–48 (2018).
40. T. Junker, V. Mohrholz, M. Schmidt, L. Siegfried, A. van der Plas, Coastal trapped wave propagation along the southwest african shelf as revealed by moored observations. *J. Phys. Oceanogr.* **49**, 851–866 (2019).
41. M.-L. Bachèlery, S. Illig, M. Rouault, Interannual Coastal Trapped Waves in the Angola-Benguela Upwelling System and Benguela Niño and Niña events. *J. Mar. Syst.* **203**, 103262 (2020).
42. R. A. I. Koungue, S. Illig, M. Rouault, Role of interannual Kelvin wave propagations in the equatorial Atlantic on the Angola Benguela Current system. *J. Geophys. Res. Oceans.* **122**, 4685–4703 (2017).
43. R. A. I. Koungue, M. Rouault, S. Illig, P. Brandt, J. Jouanno, Benguela Niños and Benguela Niñas in forced ocean simulation from 1958 to 2015. *J. Geophys. Res. Oceans.* **124**, 5923–5951 (2019).
44. M.-L. Bachèlery, S. Illig, I. Dadou, Interannual variability in the South-East Atlantic Ocean, focusing on the Benguela Upwelling System: Remote versus local forcing. *J. Geophys. Res. Oceans* **121**, 284–310 (2016).
45. N. S. Keenlyside, M. Latif, Understanding equatorial Atlantic interannual variability. *J. Climate* **20**, 131–142 (2007).
46. H. C. Nnamchi, M. Latif, N. S. Keenlyside, J. Kjellsson, I. Richter, Diabatic heating governs the seasonality of the Atlantic Niño. *Nat. Commun.* **12**, 376 (2021).
47. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
48. Y.-G. Ham, J.-H. Kim, J.-J. Luo, Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).
49. S. E. Zebiak, Air–sea interaction in the equatorial Atlantic region. *J. Climate* **6**, 1567–1586 (1993).
50. R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, P. Battaglia, Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
51. H. Zuo, M. A. Balmaseda, S. Tietsche, K. Mogensen, M. Mayer, The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Sci.* **15**, 779–808 (2019).
52. C. Buontempo, S. N. Burgess, D. Dee, B. Pinty, J.-N. Thépaut, M. Rixen, S. Almond, D. Armstrong, A. Brookshaw, A. L. Alos, B. Bell, C. Bergeron, C. Cagnazzo, E. Comyn-Platt, E. Damasio-Da-Costa, A. Guillory, H. Hersbach, A. Horányi, J. Nicolas, A. Obregon, E. P. Ramos, B. Raoult, J. Muñoz-Sabater, A. Simmons, C. Soci, M. Suttie, F. Vamborg, J. Varndell, S. Vermoote, X. Yang, J. G. de Marcilla, The Copernicus Climate Change Service: Climate science in action. *Bull. Am. Meteorol. Soc.* **103**, E2669–E2687 (2022).
53. F. Ling, J.-J. Luo, Y. Li, T. Tang, L. Bai, W. Ouyang, T. Yamagata, Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. *Nat. Commun.* **13**, 7681 (2022).
54. Y. Okumura, S.-P. Xie, Some overlooked features of tropical Atlantic climate leading to a new Niño-like phenomenon. *J. Climate* **19**, 5859–5874 (2006).
55. H. Ding, N. S. Keenlyside, M. Latif, Equatorial Atlantic interannual variability: Role of heat content. *J. Geophys. Res. Oceans.* **115**, doi.org/10.1029/2010JC006304 (2010).
56. I. C. Handoh, G. R. Bigg, A self-sustaining climate mode in the tropical Atlantic, 1995–97: Observations and modelling. *Q. J. Roy. Meteorol. Soc.* **126**, 807–821 (2000).
57. R. Seager, R. Murtugudde, N. Naik, A. Clement, N. Gordon, J. Miller, Air–sea interaction and the seasonal cycle of the subtropical anticyclones. *J. Climate* **16**, 1948–1966 (2003).
58. I. Richter, C. R. Mechoso, A. W. Robertson, What determines the position and intensity of the South Atlantic anticyclone in austral winter?—An AGCM study. *J. Climate* **21**, 214–229 (2008).
59. W. Cabos, D. V. Sein, J. G. Pinto, A. H. Fink, N. V. Koldunov, F. Alvarez, A. Izquierdo, N. Keenlyside, D. Jacob, The South Atlantic Anticyclone as a key player for the representation of the tropical Atlantic climate in coupled climate models. *Climate Dyn.* **48**, 4051–4069 (2017).

60. J. F. Lübbecke, P. Brandt, M. Dengler, R. Kopte, J. Lüdke, I. Richter, M. Sena Martins, P. C. M. Tchipalanga, Causes and evolution of the southeastern tropical Atlantic warm event in early 2016. *Climate Dynam.* **53**, 261–274 (2019).
61. L. C. Aroucha, J. F. Lübbecke, M. Körner, R. A. Imbol Koungue, F. M. Awo, The influence of freshwater input on the evolution of the 1995 Benguela Niño. *J. Geophys. Res. Oceans.* **129**, e2023JC020241 (2024).
62. S.-K. Lee, H. Lopez, F. P. Tuchen, D. Kim, G. R. Foltz, A. T. Wittenberg, On the genesis of the 2021 Atlantic Niño. *Geophys. Res. Lett.* **50**, e2023GL104452 (2023).
63. "The Atlantic Niño's role in India's erratic monsoon," *The Indian Express*, 30 September 2021; https://indianexpress.com/article/opinion/columns/atlantic-nino-imd-prediction-monsoon-forecast-7540320/.
64. M. Körner, P. Brandt, S. Illig, M. Dengler, A. Subramaniam, M.-L. Bachèlery, G. Krahmann, Coastal trapped waves and tidal mixing control primary production in the tropical Angolan upwelling system. *Sci. Adv.* **10**, eadj6686 (2024).
65. M. Gupta, H. Kodamana, S. Sandeep, Prediction of ENSO beyond spring predictability barrier using deep convolutional LSTM networks. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5 (2022).
66. B. Mu, S. Ma, S. Yuan, H. Xu, "Applying convolutional LSTM network to predict El Niño events: Transfer learning from the data of dynamical model and observation," in *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)* (IEEE, 2020), pp. 215–219.
67. N. Saeed, M. Alam, R. G. Nyberg, A multimodal deep learning approach for gravel road condition evaluation through image and audio integration. *Transp. Eng.* **16**, 100228 (2024).
68. J. Kim, M. Kwon, S.-D. Kim, J.-S. Kug, J.-G. Ryu, J. Kim, Spatiotemporal neural network with attention mechanism for El Niño forecasts. *Sci. Rep.* **12**, 7204 (2022).
69. Y. Wang, Y. Zhang, G.-G. Wang, Forecasting ENSO using convolutional LSTM network with improved attention mechanism and models recombined by genetic algorithm in CMIP5/6. *Inform. Sci.* **642**, 119106 (2023).
70. L. Zhou, R.-H. Zhang, A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Sci. Adv.* **9**, eadf2827 (2023).
71. J. T. Trok, E. A. Barnes, F. V. Davenport, N. S. Diffenbaugh, Machine learning–based extreme event attribution. *Sci. Adv.* **10**, eadl3242 (2024).
72. K. R. Patil, T. Doi, V. R. Jayanthi, S. Behera, Deep learning for skillful long-lead ENSO forecasts. *Front. Clim.* **4**, 1058677 (2023).
73. K. R. Patil, T. Doi, S. K. Behera, Predicting extreme floods and droughts in East Africa using a deep learning approach. *NPJ Clim. Atmos. Sci.* **6**, 1–8 (2023).
74. H. Lin, J. Tang, S. Wang, S. Wang, G. Dong, Deep learning downscaled high-resolution daily near surface meteorological datasets over East Asia. *Sci. Data.* **10**, 890 (2023).
75. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **128**, 336–359 (2020).
76. C. Yang, S. Masina, A. Storto, Historical ocean reanalyses (1900–2010) using different data assimilation strategies. *Q. J. Roy. Meteorol. Soc.* **143**, 479–493 (2017).
77. P. Poli, H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. H. Tan, C. Peubey, J.-N. Thépaut, Y. Trémolet, E. V. Hólm, M. Bonavita, L. Isaksen, M. Fisher, ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate* **29**, 4083–4097 (2016).
78. N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, A. Kaplan, Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmospheres* **108**, 4407 (2003).
79. S. Guinehut, A.-L. Dhomps, G. Larnicol, P.-Y. Le Traon, High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Sci.* **8**, 845–857 (2012).
80. S. Mulet, M.-H. Rio, A. Mignot, S. Guinehut, R. Morrow, A new estimate of the global 3D geostrophic ocean circulation based on satellite data and in-situ measurements. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **77–80**, 70–81 (2012).
81. M.-I. Pujol, Y. Faugère, G. Taburet, S. Dupuy, C. Pelloquin, M. Ablain, N. Picot, DUACS DT2014: The new multi-mission altimeter data set reprocessed over 20 years. *Ocean Sci.* **12**, 1067–1090 (2016).
82. G. Liu, A. Bracco, J. Brajard, Systematic bias correction in ocean mesoscale forecasting using machine learning. *J. Adv. Model. Earth Syst.* **15**, e2022MS003426 (2023).
83. E. Di Lorenzo, N. Schneider, K. M. Cobb, P. J. S. Franks, K. Chhak, A. J. Miller, J. C. McWilliams, S. J. Bograd, H. Arango, E. Curchitser, T. M. Powell, P. Rivière, North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett.* **35**, doi.org/10.1029/2007GL032838 (2008).
84. W. Ebisuzaki, A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Climate* **10**, 2147–2153 (1997).

Acknowledgments

Funding: M.-L.B. has received funding from the European Union's Horizon 2020 Research and Innovation Program for the project BENGUP under the Marie Skłodowska-Curie grant agreement ID 101025655. The study was also carried out under the EU H2020 under grant agreement 817578 (TRIATLAS project). We acknowledge support of the Bjerknes Climate Prediction Unit with funding from the Trond Mohn Foundation (grant BFS2018TMT01). **Author contributions:** Conceptualization: M.-L.B., J.B., M.P., S.I., and N.K.; data curation: M.-L.B.; formal analysis: M.-L.B.; methodology: M.-L.B., J.B., M.P., and N.K.; visualization: M.-L.B.; software: M.-L.B., J.B., and M.P.; supervision: N.K.; writing—original draft: M.-L.B; writing—review and editing: M.-L.B., J.B., M.P., S.I., and N.K.; funding acquisition: M.-L.B. and N.K.; resources: M.-L.B. and N.K.; project administration: M.-L.B. and N.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The datasets used in this paper can be downloaded from the following locations: CHORE_RL, http://c-glors.cmcc.it; ORAS5, https://cds.climate.copernicus.eu/datasets/reanalysis-oras5?tab=overview; C3S seasonal hindcasts, https://cds.climate.copernicus.eu/datasets/seasonal-monthly-single-levels?tab=overview; ARMOR3D, https://data.marine.copernicus.eu/product/MULTIOBS_GLO_PHY_TSUV_3D_MYNRT_015_012/description. All other data needed to evaluate the conclusions in this paper are present in the paper and/or the Supplementary Materials. The deep learning models were developed using standard libraries in open-source platforms including Pytorch.

Submitted 16 August 2024
Accepted 28 February 2025
Published 2 April 2025
10.1126/sciadv.ads5185

Science Advances
AAAS

Supplementary Materials for

Predicting Atlantic and Benguela Niño events with deep learning

Marie-Lou Bachèlery *et al.*

Corresponding author: Marie-Lou Bachèlery, bachelery.marielou@gmail.com

Sci. Adv. **11**, eads5185 (2025)
DOI: 10.1126/sciadv.ads5185

This PDF file includes:

Supplementary Notes S1 to S5
Figs. S1 to S11

Introduction:

These **supplementary materials** provide extra diagnostics to assess the prediction skill of the CNN model compared with the C3S seasonal prediction systems (figs. S1 to S9). They also include details of the hyperparameter tuning analysis of the CNN model.

Hyperparameter tuning of the CNN:

The targeted hyperparameter tuning focused on the Benguela Niño index (BNi), utilizing the CHORE_RL dataset (1900–2010), with a 70/30 training-validation split. The study evaluated the CNN's performance for lead times of 1 to 4 months, which represent the model's statistically significant forecast horizons. Due to computational constraints, we limited the tuning to the BNi and tested key hyperparameters through a grid search approach, including three optimizers (RMSProp, Adam, and NAdam), five learning rates (0.001 to 5e-5), and four hidden feature sizes (10, 30, 50, 70). Other hyperparameters were kept as described in the manuscript for consistency. The training process resembles the one described in the manuscript. For each hyperparameter combination (or configuration), we performed five independent training runs and averaged the correlation coefficients to reduce the internal variability of the training process. **Figure S10** illustrates these averaged results, with our settings marked by the blue star, and the best-performing configurations marked by yellow stars. NAdam achieved the highest correlation at lead times 1 (correlation of 0.83) and 2 (correlation of 0.54), although its performance was comparable to the settings used in our study (correlation of 0.82 and 0.5 for lead 1 and 2 respectively). **Figure S11** (Taylor diagram) confirms that the selected parameters yield performance closely aligned with the best configurations from the hyperparameter tuning. The averaged correlation (standard deviation) from lead 1 to 4 is 0.43 (1.07) for our settings compared to 0.46 (1.03) for the best hyperparameter combination.

1 Supplementary Note 1: All-months Root Mean Square Error (RMSE) skills assessment

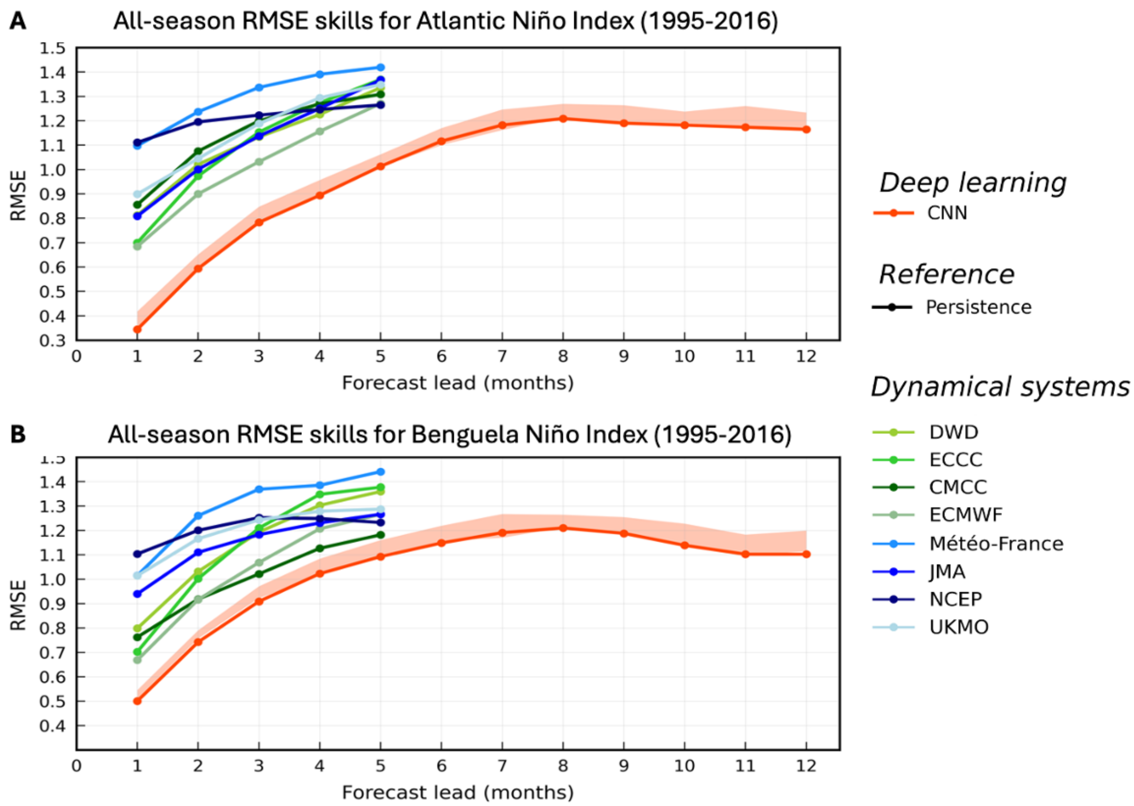

Figure S1: Root Mean Square Error (RMSE) performances of state-of-the-art C3S dynamical systems and our deep learning model. (**A**) The Atlantic Niño index [ANi: 20°W-0°E/3°S-3°N] RMSE as a function of lead time (months) for the CNN deep-learning model (orange) and the C3S prediction models (green and blue colors). The orange shading shows the highest and the lowest RMSE estimated over 20 ensemble members of the CNN model. The evaluation period is 1995-2016 and all models are compared with ORAS5. (**B**) Same as panel (A) for the Benguela Niño index [BNi: 10°S-20°S/2° coastal band].

2 Supplementary Note 2: Monthly decomposition skills of the CNN and C3S models

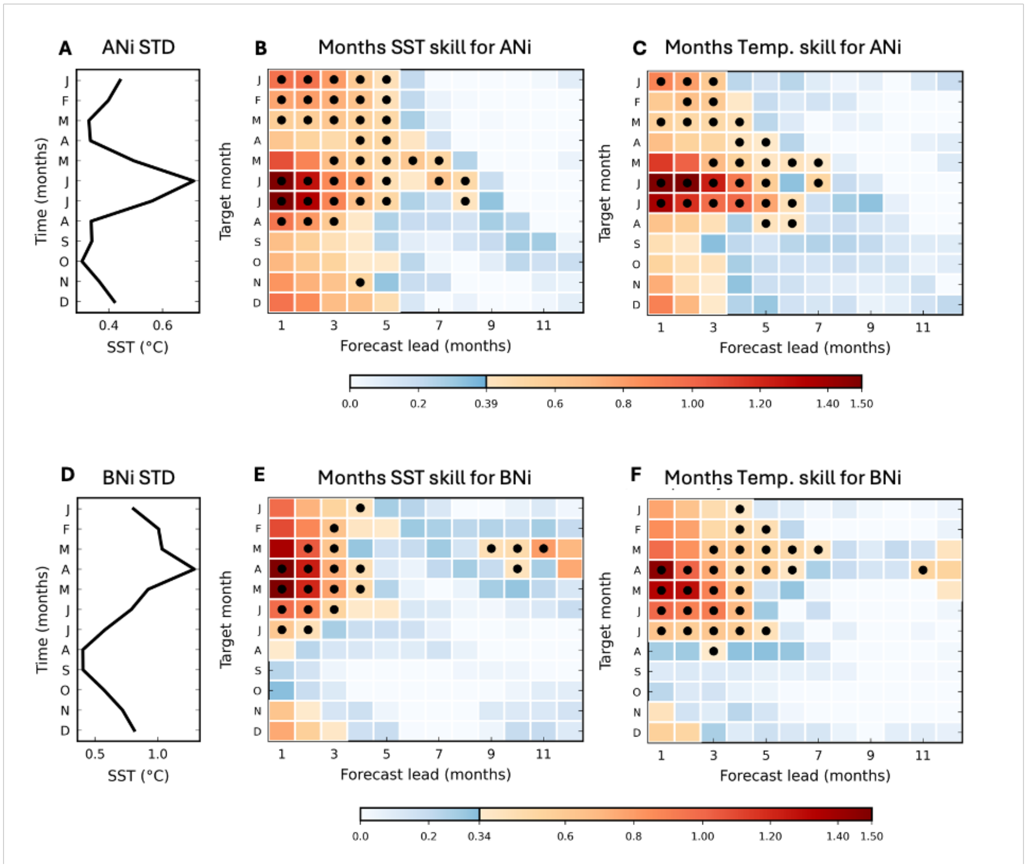

Figure S2: Variable contributions to the CNN model performances. Monthly normalized covariance skills for the Atlantic Niño (top panels; **A** to **C**) and Benguela Niño (bottom panels; **D** to **F**) indexes. (left) Monthly standard deviation of ORAS5 (°C) for ANi (**A**) and BNi (**D**). Normalized covariance decomposition (see Methods) as a function of the lead time (months) for the CNN deep-learning model with contributions from SST only (middle; **B** and **E**) and from the 100m-averaged temperature only (right; **C** and **F**). The evaluation period is 1995-2016 and all models are compared to ORAS5. Statistically significant values (>99%, estimated using a red noise significance method(86); see Methods) are indicated by the red color bar. Monthly significant correlation values above persistence are marked by a black point.

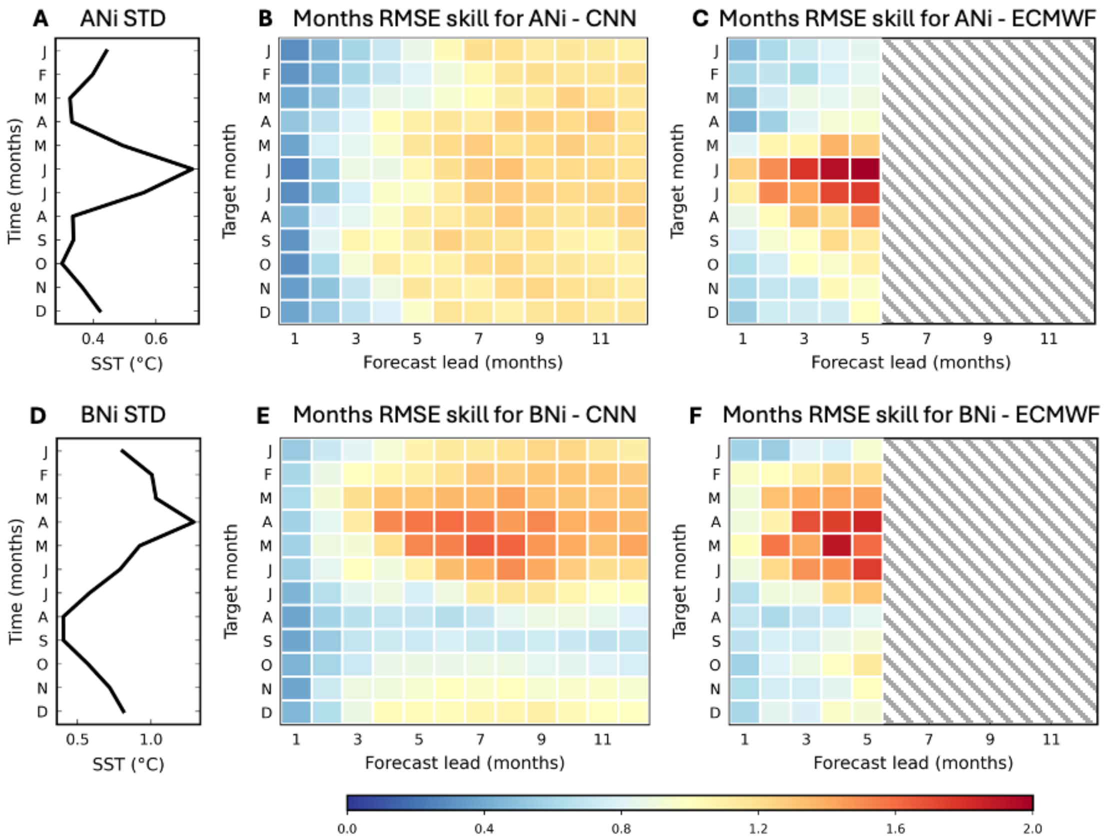

Figure S3: Monthly performances for CNN and ECMWF models. Monthly Root Mean Square Error (RMSE) decomposition skills for the Atlantic Niño (top panels; **A** to **C**) and Benguela Niño (bottom panels; **D** to **F**) indexes. (left) Monthly standard deviation of ORAS5 (°C) for the ANi (**A**) and BNi (**D**). Seasonal RMSE decomposition (see Methods) as a function of the lead time (months) for the CNN deep-learning model (middle; **B** and **E**) and the ECMWF dynamical model (right; **C** and **F**). The evaluation period is 1995-2016 and all models are compared to ORAS5.

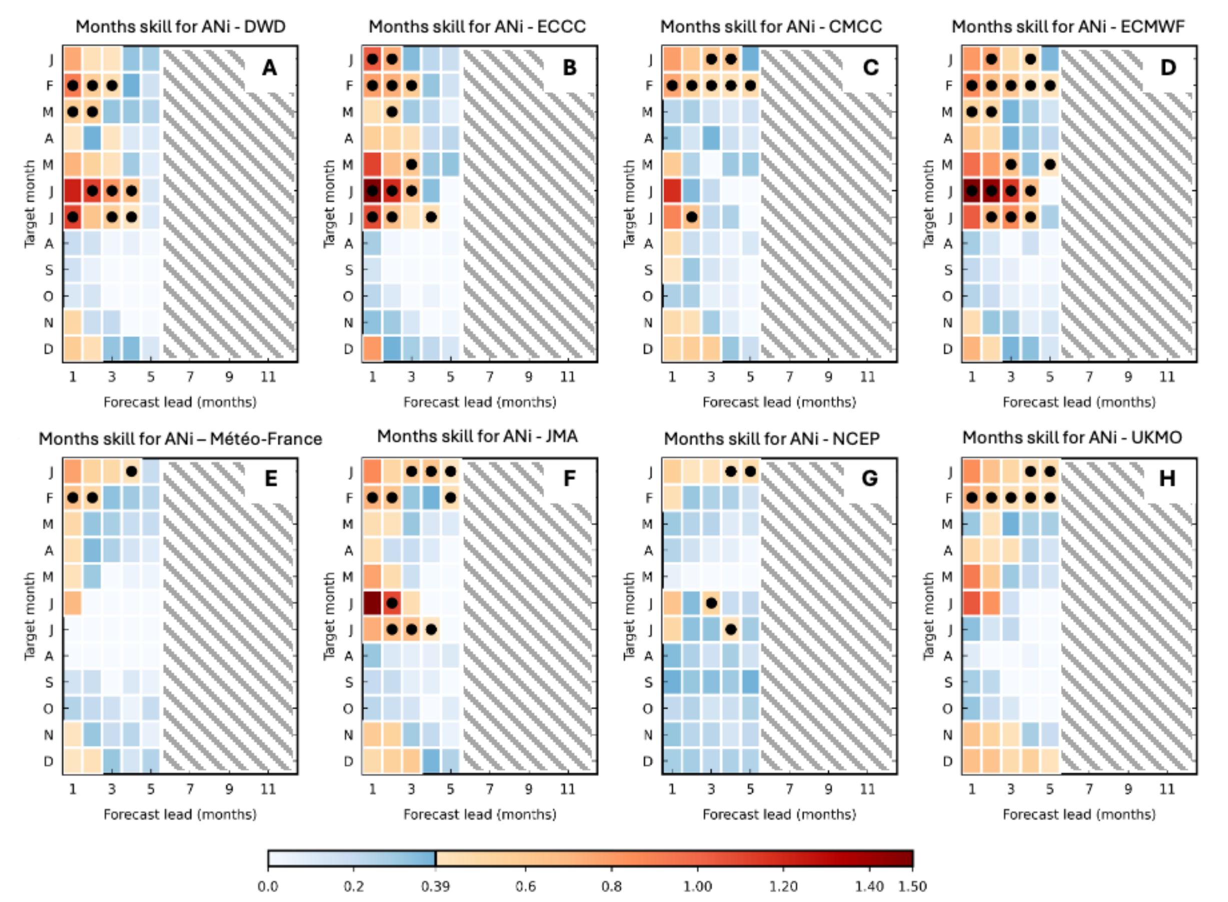

Figure S4: Monthly normalized covariance skills of C3S for the ANi. Normalized covariance decomposition (see Methods) as a function of the lead time (months) for (**A**) DWD, (**B**) ECCC, (**C**) CMCC, (**D**) ECMWF, (**E**) Météo-France, (**F**) JMA, (**G**) NCEP and (**H**) UKMO. The evaluation period is 1995-2016 and all models are compared to ORAS5. Statistically significant values (>99%, estimated using a red noise significance method; see Methods) are indicated by the red color bar. Correlations above persistence are shown by the black points.

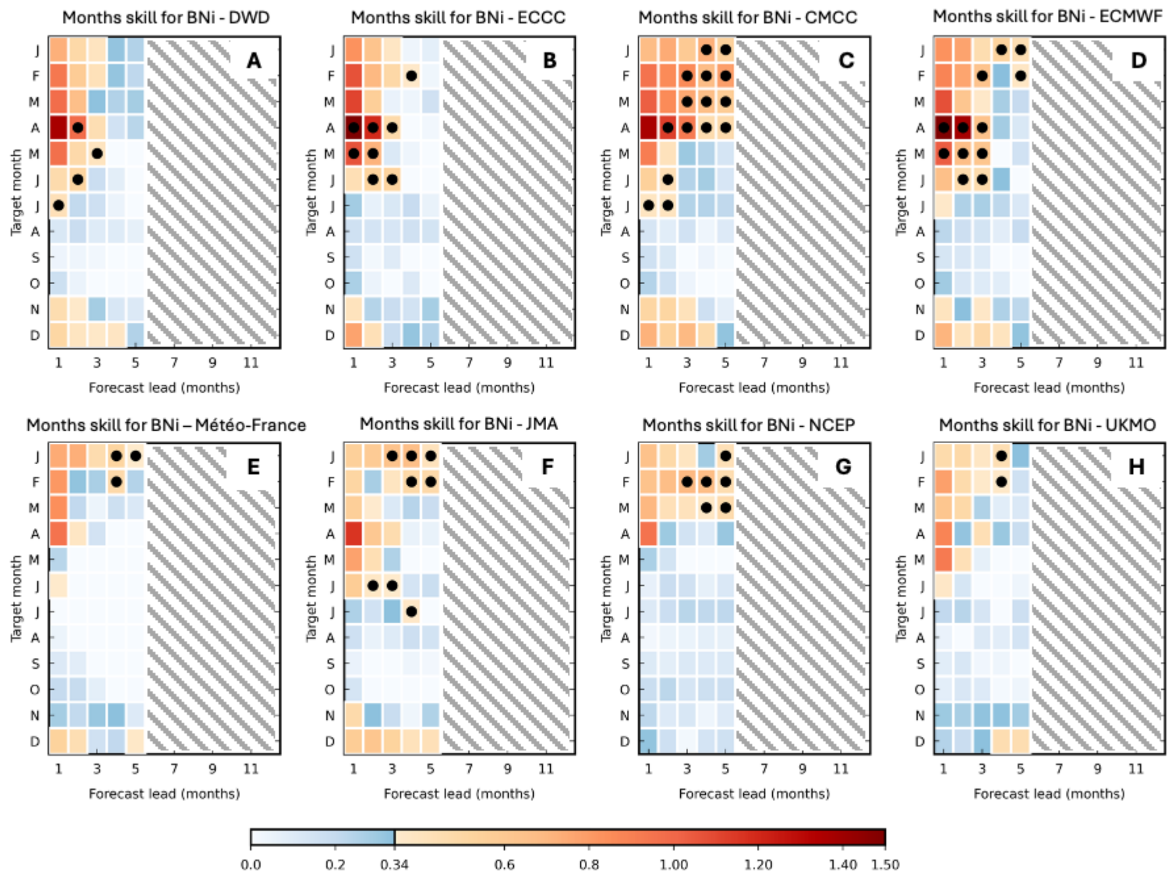

Figure S5: Monthly normalized covariance skills of C3S for the BNi. (**A** to **H**) Same as fig. S4 for the BNi.

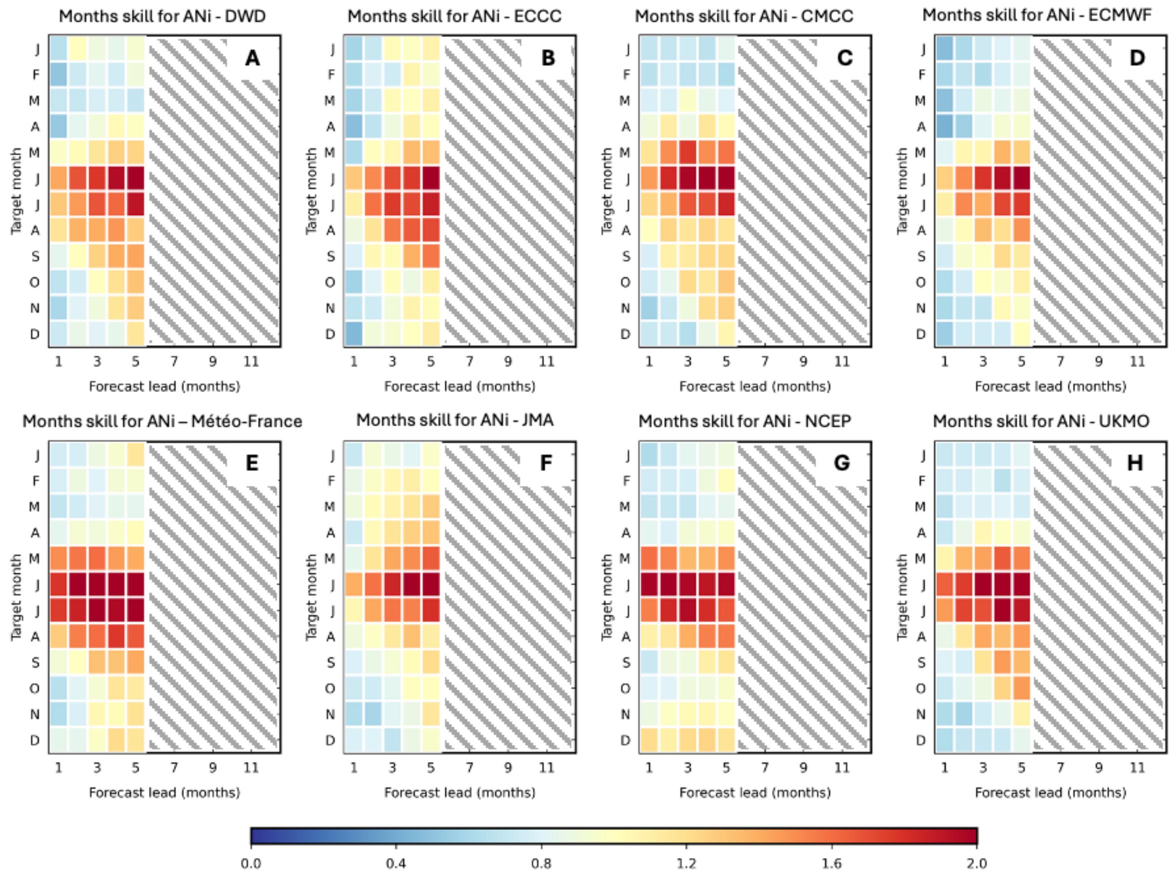

Figure S6: Monthly root mean square error decomposition skills of C3S for the ANi. Seasonal RMSE decomposition (see Methods) as a function of the lead time (months) for (**A**) DWD, (**B**) ECCC, (**C**) CMCC, (**D**) ECMWF, (**E**) Météo-France, (**F**) JMA, (**G**) NCEP and (**H**) UKMO. The evaluation period is 1995-2016 and all models are compared to ORAS5.

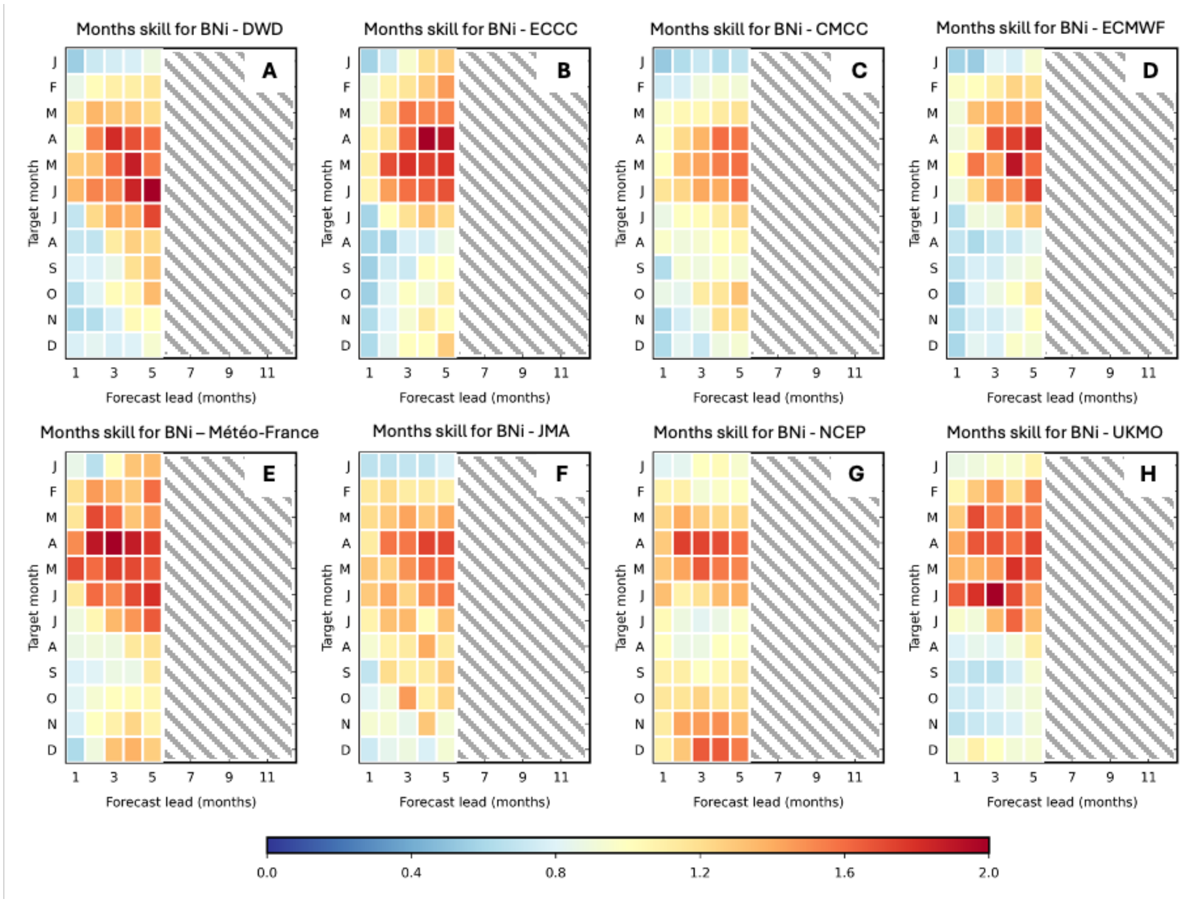

Figure S7: Monthly root mean square error decomposition skills of C3S for the BNi. (**A** to **H**) Same as fig. S6 but for BNi.

3 Supplementary Note 3: Benguela Niño/Niña events detection skills

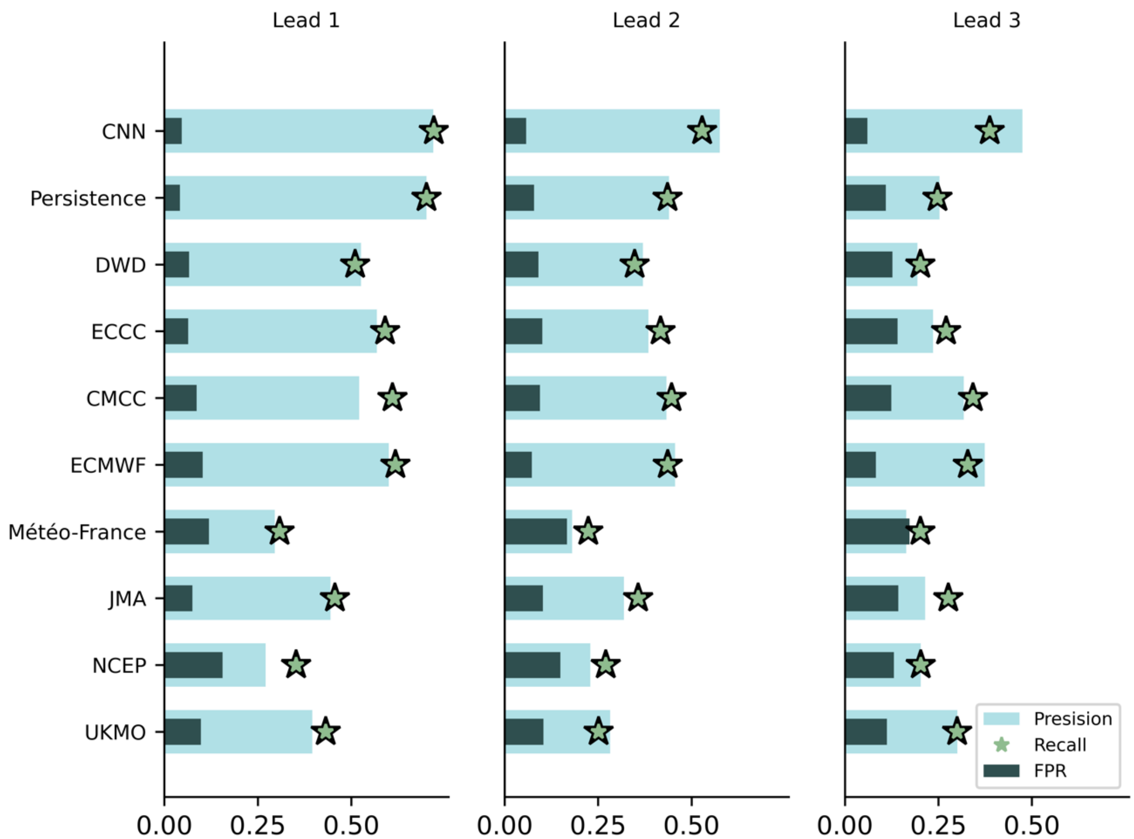

Figure S8: **Ability to predict the occurrence of Benguela Niños/Niñas**. Precision (light blue bars), recall (stars) and False Positive Rate (FPR; dark green bars; see Methods) for Benguela Niño/Niña events combined, using the CNN model, the persistence (ORAS5), and the C3S dynamical prediction systems. Scores are evaluated against ORAS5 for leads 1 (left panel) to 3 (right panel). Events are selected when the BNi amplitude exceeds 1 standard deviation for two consecutive months.

4 Supplementary Note 4: Benguela Niño/Niña events heatmaps

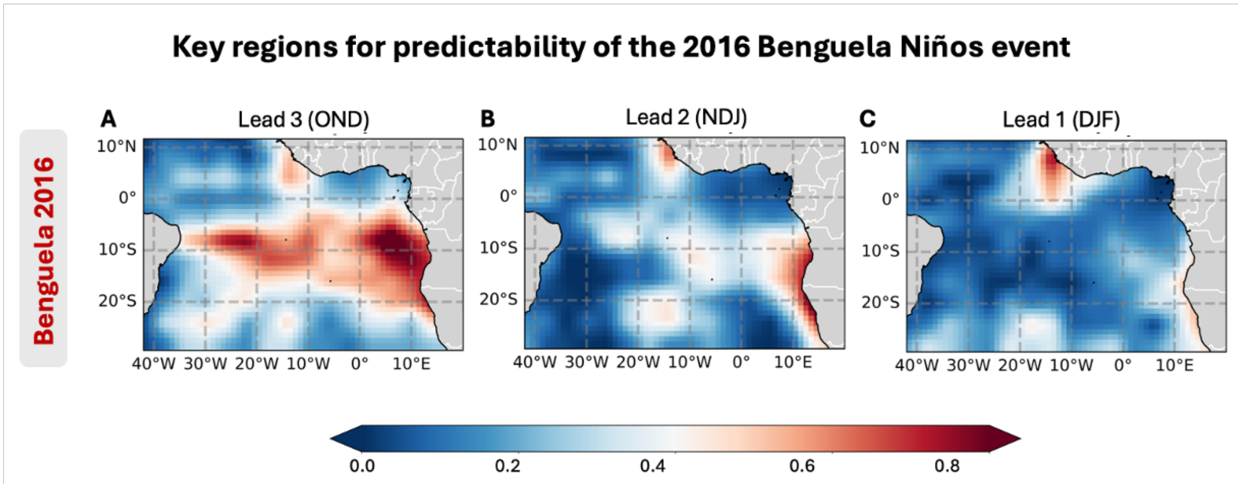

Figure S9: CNN Predictors Hotspots for the march 2016 Benguela Niño forecasts. Heatmaps at (**A**) lead 3 (i.e. using precursors from October-November-December), (**B**) lead 2 (i.e. with precursors from November-December-January) and (**C**) lead 1 (i.e. with precursors from December-January-February).

5 Supplementary Note 5: Hyperparameters tuning analysis

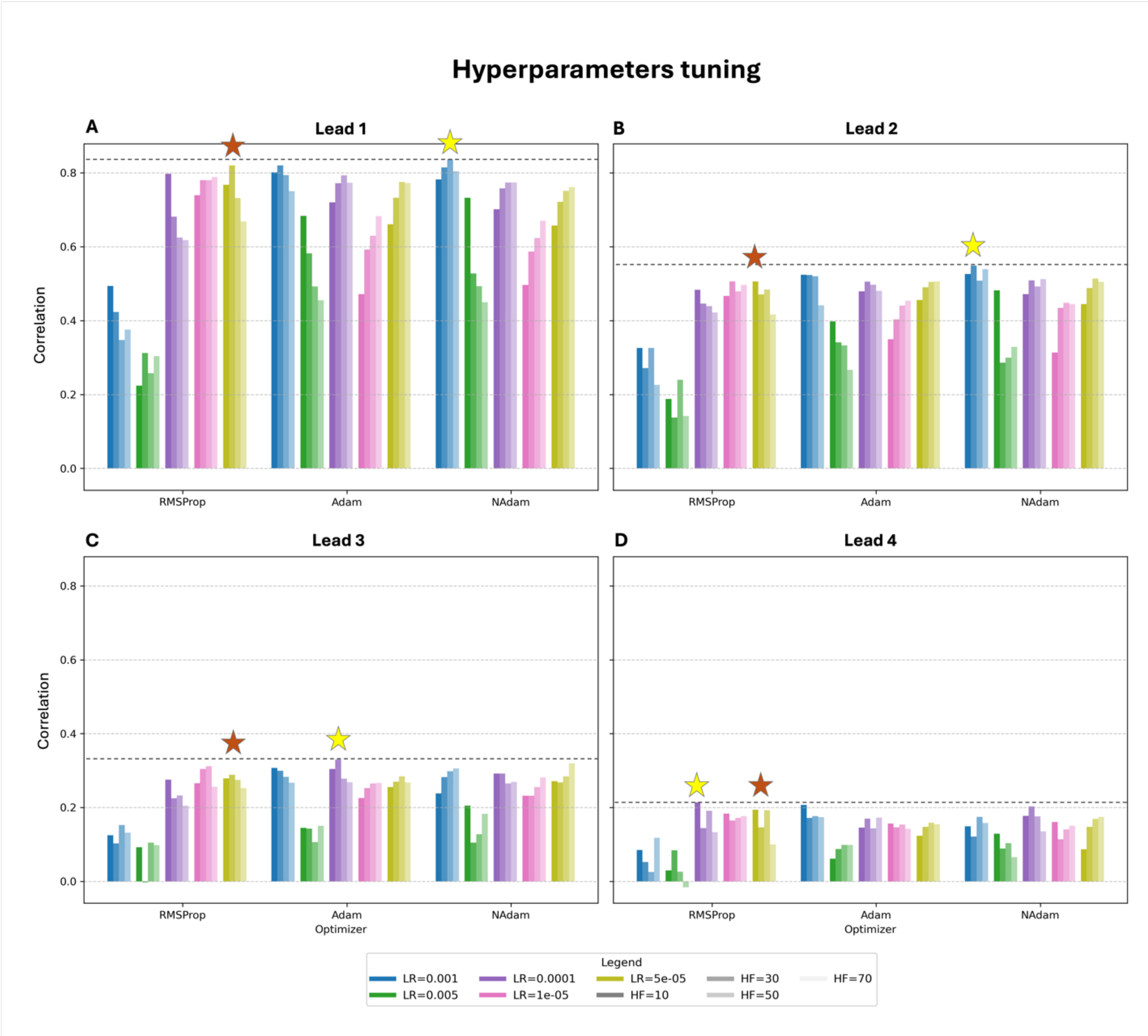

Figure S10: CNN hyperparameter tuning for the Benguela Niño index (BNi). All-month correlation coefficients at lead times (**A**) 1 month, (**B**) 2 months, (**C**) 3 months, and (**D**) 4 months are shown for three optimizers: RMSProp, Adam, and NAdam. The study includes five learning rates (LR): 0.001 (blue bars), 0.005 (green bars), 0.0001 (purple bars), 1e-5 (pink bars), and 5e-5 (yellow bars), as well as four hidden feature sizes (HF: 10, 30, 50, and 70), represented by varying bar transparency from opaque to more transparent. The validation set spans from 1966 to 1992 using the CHORE_RL reanalysis datasets. Correlation values corresponding to the hyperparameter settings used in the manuscript are marked with a red star. The best correlation values are highlighted with a yellow star and indicated by the horizontal dark grey dashed line.

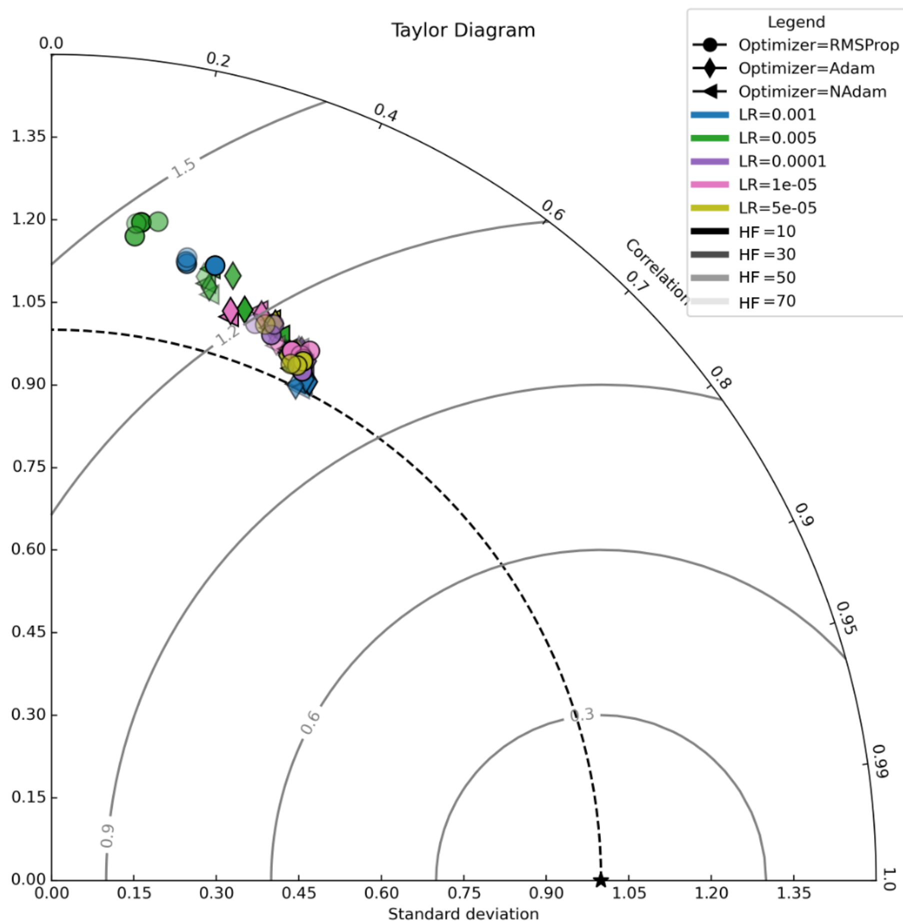

Figure S11: Hyperparameter tuning for the Benguela Niño index (BNi). Normalized Taylor diagram summarizing the performance of the CNN model in predicting BNi, with different hyperparameters combinations: optimizers (symbols), learning rates (LR: colors) and number of hidden features (HF: transparency). All-months correlation coefficients and standard deviation are averaged over lead times from 1 to 4 months. The validation set spans from 1966 to 1992 using the CHORE_RL reanalysis datasets.