Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL086423

# Probabilistic Forecasting of El Niño Using Neural Network Models

**Paul Johannes Petersik**[1] **and Henk A. Dijkstra**[1,2]

[1]Institute for Marine and Atmospheric Research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, The Netherlands, [2]Centre for Complex Systems Studies (CCSS), Department of Physics, Utrecht University, Utrecht, The Netherlands

**Key Points:**
- Gaussian density and quantile regression neural networks are used to make probabilistic forecasts of the ENSO
- High-correlation skill is found for long lead times for the period between 1963 and 2017
- The estimation of the predictive uncertainty is improved with respect to a standard method for the period between 1982 and 2017

**Supporting Information:**
- Supporting Information S1

**Correspondence to:**
P. J. Petersik,
paul.j.petersik@gmail.com

**Citation:**
Petersik, P. J., & Dijkstra, H. A. (2020). Probabilistic forecasting of El Niño using neural network models. *Geophysical Research Letters*, *47*, e2019GL086423. https://doi.org/10.1029/2019GL086423

Received 6 DEC 2019
Accepted 9 MAR 2020
Accepted article online 13 MAR 2020



**Abstract** We apply Gaussian density neural network and quantile regression neural network ensembles to predict the El Niño–Southern Oscillation. Both models are able to assess the predictive uncertainty of the forecast by predicting a Gaussian distribution and the quantiles of the forecasts, respectively. This direct estimation of the predictive uncertainty for each given forecast is a novel feature in the prediction of the El Niño–Southern Oscillation by statistical models. The predicted mean and median, respectively, show a high-correlation skill for long lead times ($r = 0.5$, 12 months) for the 1963–2017 evaluation period. For the 1982–2017 evaluation period, the probabilistic forecasts by the Gaussian density neural network can better estimate the predictive uncertainty than a standard method to assess the predictive uncertainty of statistical models.

**Plain Language Summary** We apply, for the first time, machine learning models that can directly estimate the uncertainty of the given forecasts of the El Niño phenomenon. Usually, machine learning models for the prediction of the El Niño phenomenon only forecast one value of the so-called Oceanic Niño Index. In contrast, our models predict which values are likely to be observed and which are not. We find that the models have high-correlation skill for long lead times for an evaluation between 1963 and 2017. Moreover, the estimation of the predictive uncertainty is superior to simpler methods for an evaluation between 1982 and 2017.

## 1. Introduction

The interannual climate variability of the tropical Pacific is strongly dominated by the El Niño–Southern Oscillation (ENSO). Two characteristics make El Niño forecasting very interesting for the climate research community. First, the predictability horizon of El Niño forecasts is at least a factor ten larger than the one of weather forecasts. Because of the strong autocorrelation of ENSO dynamics and the presence of a memory component, it is (sometimes) possible to make skillful forecasts for a year ahead (e.g., Balmaseda et al., 1995). Second, ENSO has distinct influences on the climate around the globe through well-known teleconnections (e.g., Diaz et al., 2001). The combination of a relatively large predictability horizon and global climate influence makes ENSO forecasting a very worthwhile field of research.

There have been many reviews on El Niño predictability (Barnston et al., 2019, 2012; Chen & Cane, 2008; Latif, 1998; Tang et al., 2018; Timmermann et al., 2018), indicating that both statistical models (those capturing behavior of past statistics) and dynamical models (i.e., those based on underlying physical conservation laws) are used for El Niño prediction. Multimodel ensemble results are given at the International Research Institute for Climate and Society (https://iri.columbia.edu/our-expertise/climate/forecasts/#ENSO&urluscore;Forecasts). The National Centers for Environmental Prediction Climate Forecast System CFSv2 (Saha et al., 2014) gives a dynamical single-model ensemble forecast. Moreover, a superensemble consisting out of multiple models recently has been developed in China to provide deterministic and probabilistic ENSO forecasts (Tang et al., 2018).

Machine learning methods, in particular Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) are powerful statistical models that have been applied to El Niño prediction. CNNs trained on CMIP5 data and on reanalysis showed a better forecasting skill than most dynamical models and, in particular, forecast skill remained high up to lead times of about 17 months (Ham et al., 2019). A recent review by Dijkstra et al. (2019) provides an overview of these approaches and of the particular methods used.

Early work (e.g., Tangang et al., 1997; 1998; Wu et al., 2006) used feed-forward ANNs with PCs of the leading EOFs of wind stress or sea level pressure as input. For the 1980s and the 1990s these models had a remarkable skill for long lead times in predicting the Oceanic Niño Index (ONI; $r > 0.6$ for the 12-month lead time). However, prediction skill was considerably weaker for the 1950s, 1960s, and the 1970s. Moreover, one of these models dramatically underperformed in a later study by Barnston et al. (2012) for the period 2002–2011 ($r \approx 0.0$, 7-month lead time).

More recently, Nooteboom et al. (2018) used a combination of an autoregressive integrated moving average model and an ANN to predict the ONI based on a small set of predictor variables, namely, the ONI itself, the warm water volume (WWV) of the tropical Pacific, the seasonal cycle represented by a sinusoid with a period of 1 year, and the variable $c_2$, which refers to the number of nodes in an evolving complex network (ECN), which are part of a cluster of size 2. Their model was evaluated for a test period between 2007 and 2013. Interestingly, this model showed a signal for the 2009/2010 El Niño for the 12-month lead time that was poorly predicted by other models.

Whereas dynamical models, for example, the CFSv2 (Saha et al., 2014), can give an estimation of the predictive uncertainty by making ensemble forecasts, the statistical models for the El Niño prediction lack the ability to do so. Only very recently, McDermott and Wikle (2019) used a machine learning model, a so-called Bayesian neural network, which is able to give an estimation of the predictive uncertainty. However, they just evaluated their model for a very short time period between 2015 and 2016. A comprehensive evaluation of machine learning models predicting El Niño with an estimation of the predictive uncertainty is still lacking.

Lakshminarayanan et al. (2017) proposed an ANN model, a so-called Deep Ensemble (DE), which can estimate the predictive uncertainty but is conceptually simpler than a Bayesian neural network. A DE consists out of multiple Multilayer Perceptrons, and instead of just predicting a single value, each member of the ensemble predicts the mean and the standard deviation of a Gaussian distribution. In this study, the DE approach is, for the first time, applied to and evaluated for El Niño prediction. However, here we call the model Gaussian Density Neural Network (GDNN) ensemble since the term *deep* typically refers to an ANN with at least three hidden layers.

Another method that we use in this study to estimate the predictive uncertainty of ENSO forecasts are so-called Quantile Regression Neural Networks (QRNNs). These are introduced in Taylor (2000) and, for instance, used for precipitation downscaling in Cannon (2011). A typical QRNN predicts the location of a quantile of interest. Hence, for the prediction of multiple quantiles, one needs to train one model for each quantile. However, it is as well possible to construct a QRNN such that all quantiles are predicted by one model with constraints that prevent the issue of quantile crossing, which can arise if one QRNN is trained for each quantile (e.g., Cannon, 2018). In contrast to the GDNN, the QRNN models do not put any prior assumptions onto the distribution of the predictive uncertainty.

## 2. Data and Methods

El Niño is quantified in this study using the ONI, the 3-month running-mean of the area-averaged sea surface temperature anomaly (SSTA) in the NINO3.4 region (−5–5°N, 170–120°W). Here, the SSTA values come from the ERSSTv5 data set (Huang et al., 2017). The value of the ONI is assigned to the last month of the considered season, for example, the value of the season December-January-February is assigned to February. This is done to make sure that the prediction scheme is not accidentally using any data from the future when the ONI is used as a lagged predictor.

### 2.1. Predictor Variables

The ONI is the target variable which is predicted based on a set of predictor variables. The first predictor variable for our model is the ONI itself. This is because the ONI has a strong autocorrelation and hence already caries information about its future state.

The second predictor is the WWV, the volume of water above the 20 °C isotherm in the tropical Pacific (5°S to 5°N, 120–80°W), which is tightly connected to the recharge oscillator theory of the ENSO (Jin, 1997). It can be a strong predictor for upcoming El Niño/La Niña events but appears to have a changing predictive skill over time (Bunge & Clarke, 2014). Time series of the WWV are obtained from the Bureau National Operations Centre (https://www.pmel.noaa.gov/elnino/upper-ocean-heat-content-and-enso) for

the period between January 1980 and December 2017. For the time between January 1960 and December 1979, we use the WWV proxy which was introduced in Bunge and Clarke (2014). For simplicity, we refer to WWV in the following as the combined time series of WWV and WWV proxy.

The third predictor variable for our model is the Dipole Mode Index (DMI) of the Indian Ocean Dipole. The DMI is the difference between the monthly area-averaged SSTA of the western (10°S to 10°N, 50–70°E) and southeastern (10°S to 0°N, 90–110°E) equatorial Indian Ocean. Time series of the DMI are from the Earth System Research Laboratory's Physical Sciences Division (https://www.esrl.noaa.gov/psd/gcos&urluscore; wgsp/Timeseries/). In past studies, the DMI was found to have a predictive skill on the ONI (Izumo et al., 2010; Wieners et al., 2016) as well as an influence on the spatial distribution of the Pacific SSTAs during an El Niño event (Hameed et al., 2018).

A further predictor variable, denoted $\bar{\tau}_{x,\mathrm{WP}}$, for our model is the area-average of a quantity which is proportional to the zonal wind stress anomaly in the western Pacific (2.5°N to 2.5°S, 120°E to 160°W). It is calculated from the zonal wind, $u$, and the meridional wind, $v$, components from the National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis data set (Kalnay et al., 1996) on a 2.5° × 2.5° grid by

$$\bar{\tau}_{x,\mathrm{WP}} = |\vec{U}|u \tag{1}$$

with $\vec{U} = (u, v)$. Westerly wind anomalies excite equatorial Kelvin and Rossby waves in the ocean, which can lead to the development of an El Niño/La Niña by the delayed oscillator mechanism (Suarez & Schopf, 1988). Furthermore, westerly wind anomalies can induce a convective ocean-atmosphere instability (Clarke, 2014) due to which warm water from the warm pool in the western Pacific (WP) is pushed further and further to the east in case of a developing El Niño.

The fifth predictor variable is a metric of an ECN which is computed for the monthly sea surface height anomaly from the ORAS4 data set (Balmaseda et al., 2013). Some network metrics contain early warning signals for upcoming El Niño/La Niña events for very long lead times more than a year ahead (Ludescher et al., 2014; Rodríguez-Méndez et al., 2016). To compute an ECN, successive complex networks are computed for a shifting time window (Radebach et al., 2013). In this study, the time window spans one year and is shifted by 1 month for a new time step. We determine an ECN from the sea surface height anomaly field in the tropical Pacific (30°N to 30°S). In this study we use the corrected Hamming distance, $\mathcal{H}^*$, as a predictor. The corrected Hamming distance is a variable which displays how strong the network changed from the last time step to the current one. We assume that a strong change in the connectivity of the network indicates that the network undergoes a percolation transition, which in turn can be linked to the occurrence of ENSO events (Rodríguez-Méndez et al., 2016). More details on this network metric can be found in section SA of the supporting information (SI).

Lastly, a cosine function with a period of a year is included to incorporate information about the seasonal cycle.

Before the training of our forecast model, each predictor variable is normalized by first subtracting its mean and then dividing this difference by the standard deviation of the predictor variable. The mean and standard deviation are computed from the entire time series of the considered variable (1960–2017). Additionally, for each predictor we include the values from the seasons 3 and 6 months before the latest observed season into our predictor set. Hence, the predictor set consists out of $6 \times 3 = 18$ values. Note that, because the cosine function is included with lag values, the seasonal cycle is unambiguously represented. In the following, we refer to this set with the vector $\mathbf{x}$. For a brief overview, Figure S1 shows the time series of all considered predictor variables between 1960 and 2017.

### 2.2. Neural Network Models

GDNN and QRNN ensembles are used to predict the ONI for various lead times. Both models consist out of multiple Multilayer Perceptrons that predict either both the mean and the standard deviation of a Gaussian distribution or the locations of the quantiles, respectively. For the QRNN, all quantiles are predicted by one model where the output layer is constructed in such a way that quantile crossing is by definition not possible.

Since there is a relatively low amount of data available to train the ensembles, that is, 696 months, it is necessary to regularize the neural networks. For this, Dropout is applied on the hidden layers of the neural networks (Srivastava et al., 2014). Moreover, early stopping (Morgan & Bourlard, 1990; Prechelt, 1998)

is used in case the loss on a second so-called validation data set increases for several epochs. The specific model architectures and the training setup can be found in section SB of the SI.

A third so-called test data set is finally used to evaluate the model. The test data set is hold out during the entire training process to serve as an independent data set for the evaluation of the model. Here, we split the entire time series into six test periods, namely, 1963–1971, 1972–1981, 1982–1991, 1992–2001, 2002–2011, and 2012–2017. For each test period, we train an ensemble on the remaining time series. This time series is further split into the training and the validation data sets (Figure S3) to fit multiple ensemble members and check for overfitting (more details in section SC of the SI). We train ensembles for the 0-, 3-, 6-, 9-, 12-, 15-, 18-, and 21-month lead time.

For the evaluation of the models, we use the anomaly correlation coefficient (ACC). The ACC is the Pearson correlation coefficient between the predicted mean/median for the GDNN/QRNN ensemble and the observed ONI. Moreover, to access the probabilistic forecast skill, the quantile skill score (QSS, Bentzien & Friederichs, 2014) is used. To compute the QSS, one first computes the quantile score (QS, Friederichs & Hense, 2007;Koenker & Machado, 1999) for each quantile:

$$\mathrm{QS}(\hat{y}, y, q) = \sum_{i=1}^{N} \max(\{q(y_i - \hat{y}_i), (1-q)(y_i - \hat{y}_i)\}) \tag{2}$$

Here, $\hat{y}$ is the time series of the predicted quantile, $y$ the time series of the observed ONI, and $q \in (0, 1)$ the quantile of interest. The variable $N$ denotes the total number of observations and $i$ the $i$th prediction/observation. Now, the QSS of a quantile can be computed as follows

$$\mathrm{QSS} = 1 - \frac{\mathrm{QS}}{\mathrm{QS}_{\mathrm{ref}}} \tag{3}$$

where $\mathrm{QS}_{\mathrm{ref}}$ is the quantile score of another prediction that is considered as reference. A positive QSS indicates that the model is able to better estimate the location of the quantile than the reference. For the reference, we choose a quantity ($\sigma$) that uses the predicted mean/median from our models but estimates the predictive uncertainty following equation (5) in Barnston et al. (2015). Hence, the predictive uncertainty is estimated with a Gaussian distribution around the predicted mean/median with a standard deviation, that is, $\sigma$, computed by

$$\sigma(m, \Delta t) = \sigma_{\mathrm{clim}}(m)\sqrt{1 - (\mathrm{ACC} \quad (\Delta t))^2} \tag{4}$$

with $\sigma_{\mathrm{clim}}$ the climatological standard deviation of the ONI in the predicted target season, $m$, with a value of the ACC for the corresponding lead time, $\Delta t$. For the evaluation against the prediction of the GDNN and the QRNN ensembles, $\sigma_{\mathrm{clim}}$ and ACC are computed from the time series excluding the test decade on which this method is applied for the reference. This is done to strictly keep the test decade as an independent data set.

## 3. Results

The hindcasts on the test periods are concatenated in a successive order for the evaluation of the proposed models. For the 9-month lead time, the concatenated hindcast is shown in Figure 1 for the GDNN (a) and the QRNN (b). Both models have clear signals for El Niño as well as La Niña events with a varying strength of the predictive uncertainty. A seasonal cycle in the predictive uncertainty is present for the GDNN ensemble hindcast, whereas the predictive uncertainty of QRNN ensemble appears more irregular. A plot including the hindcasts for all considered lead times is shown in Figure S4 for the GDNN and in Figure S5 for the QRNN ensemble. Whereas the GDNN ensemble can successfully incorporate about 68% of the observed ONI values between the 16% and the 84% quantiles, the fraction of values that lie between the 16% and the 84% quantiles is systematically lower than expected for the QRNN ensembles (55.5–62.1%). The reliability diagram for the GDNN in Figure S6a reveals that the observed frequency of the predicted quantiles closely follows the perfect reliability. In contrast for the QRNN ensemble, the reliability diagram in Figure S6b shows that too many observations lie below the 2.5% and 16% quantiles and too little below the 86% and the 97.5% quantiles.

The all season ACC skills are shown in Figure 2 for the GDNN (a) and the QRNN ensembles (b). Both models have nearly identical ACC skills in the considered test decades. During 1982–2001, the ACC stays above 0.7

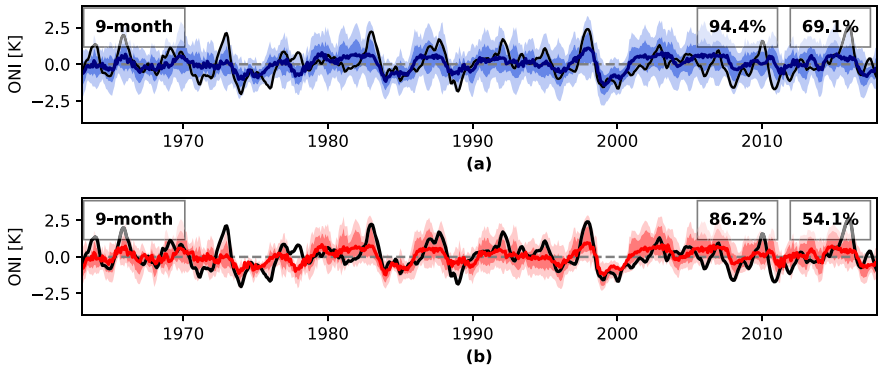


**Figure 1.** The 9-month lead time hindcast for the GDNN (a) and the (b) QRNN ensemble. The solid line indicates the predicted mean or the predicted median, respectively. The darker shading displays the area between the 16% and the 84% quantiles, the lower lighter shadings indicates the area between the 2.5% and the 16% quantiles, and the upper lighter shading the area between the 84% and 97.5% quantiles. The percentage of the observations that lie between the 2.5% and the 97.5% quantiles (left number) as well as between 16% and the 84% quantiles (right number) is shown in the upper right corner.

for lead times up to 12 months ahead. In contrast, for other test periods the ACC drops toward 0.4–0.6 for the 6-month lead time. For longer lead times in these decades, the ACC drops even further.

The seasonal ACC skills are shown in Figure 3 for the GDNN (a) and the QRNN (b). Again, the skills appear to be nearly identical between the two models. Figures 3c and 3d show the difference in the seasonal skill for the GDNN in El Niño-like periods (1982–2001) and La Niña-like periods (1963–1981 and 2002–2017), respectively. In El Niño-like periods, the ACC does not have a strong seasonal variation. In contrast for La Niña-like periods, target seasons in boreal summer do not have a significant ACC skill even for relatively short lead times such as the 3- or 6-month lead time. However, for target seasons in boreal winter in La Niña-like periods, ACC skills of about 0.5 are found up to the 12-month lead time. Similar behavior is found for the QRNN (Figure S7).

The QSS skills between 1963–2017 are shown in Figure 4 for the GDNN (a) and the QRNN (b). The plots indicate that the GDNN is only for some quantiles and lead times able to achieve a positive QSS. The QSS values for the QRNN are mostly negative. However, if one restricts the evaluation to the period between 1982 and 2017, the GDNN has mostly positive QSS values (Figure 4c). In contrast, the majority of QSS values still remain negative for QRNN. Additionally, for the GDNN ensemble and the reference method, it is possible to use the loss function of the GDNNs, the negative log-likelihood (NLL, see equation (S2) in the SI) of a Gaussian distribution, as probabilistic evaluation metric. Figure S8 shows the NLL score for the GDNN ensemble and the reference method for various lead times. Again, the GDNN ensemble has a similar skill as the reference method for the period between 1963 and 2017 but can clearly improve upon the reference method when the evaluation period is restricted to 1982–2017.

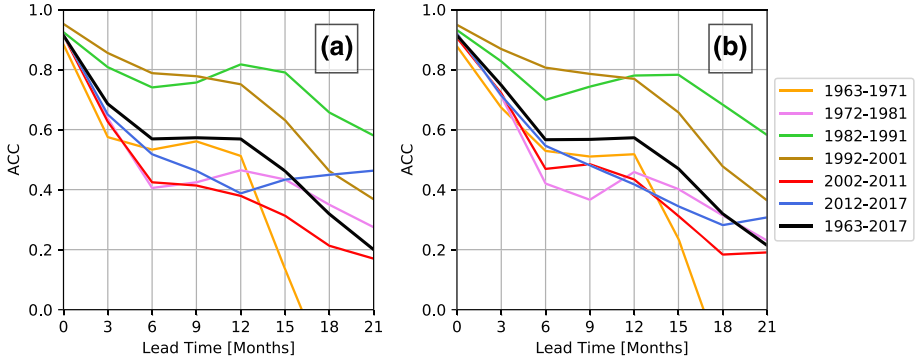


**Figure 2.** All season ACC skill of the GDNN (a) and the QRNN (b) ensemble for various lead times and decades.

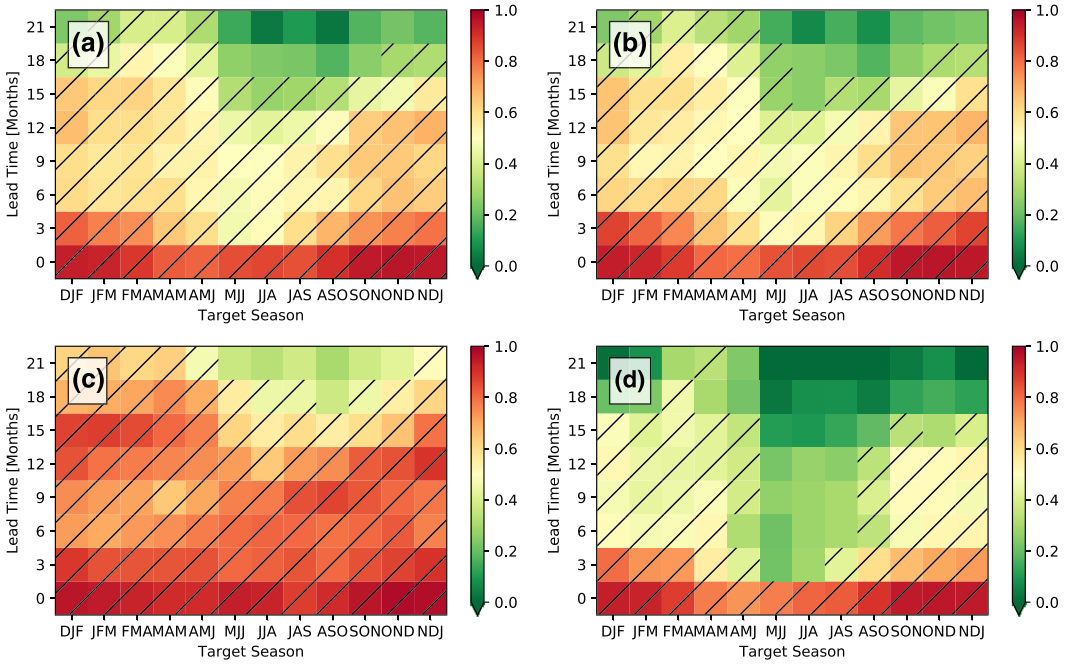


**Figure 3.** Seasonal ACC skill of the GDNN (a) and the (b) QRNN ensemble for various lead times and decades. Hatches represent an ACC skill that is significant with a 95% confidence. (c, d) Same as in (a) above for the GDNN ensemble but now (c) during El Niño-like period (1982–2001) and (d) during La Niña-like period (1963–1981 and 2002–2017).

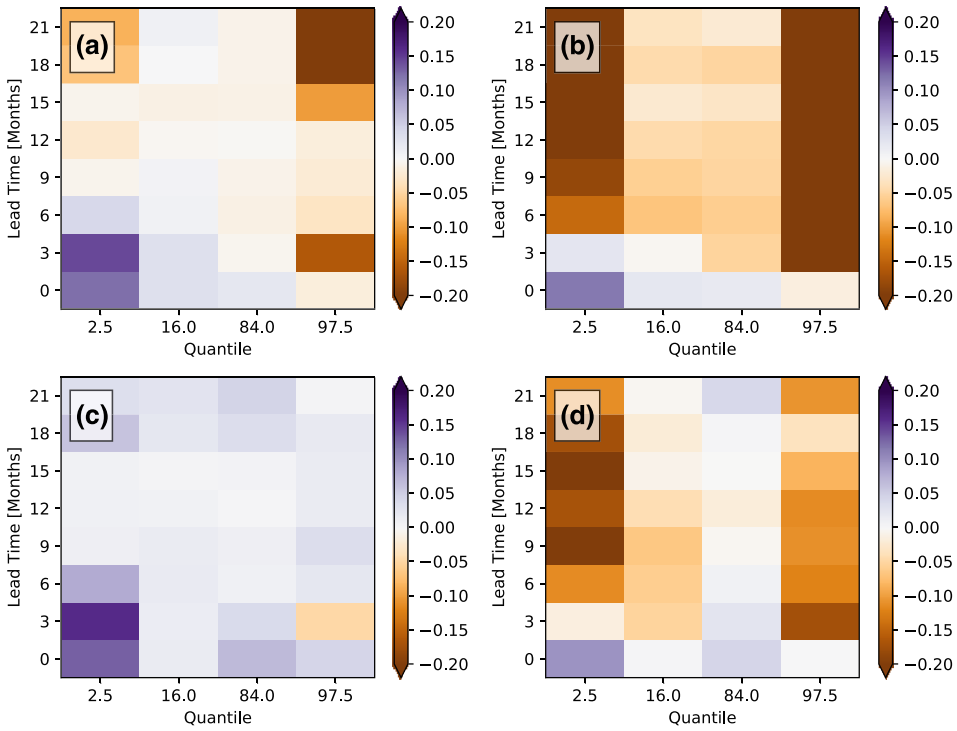


**Figure 4.** All season QSS skill of the GDNN (a, c) and the QRNN (b, d) ensemble for various lead times and quantiles for the evaluation period between 1963–2017 (a, b) and 1982–2017 (c, d).

## 4. Discussion and Conclusions

This study investigated, for the first time, the application of neural networks for the probabilistic forecasting of the ENSO. The proposed models could achieve high ACC skills for long lead times. These ACC skills are comparable to the ACC skills reported for the CNN in Ham et al. (2019). For this, compare Figure 2a in Ham et al. (2019) and Figure S9. The ACC of Ham et al. falls below 0.5 for the 17-month lead time with respect to an evaluation period between 1984 and 2017. The ACC skill of the GDNN crosses this threshold around the 16.5-month lead time. Here, lead time is defined as in Ham et al. (2019) being the time that passes between end of last observed data to middle of the 3-month target season. This result is remarkable because the predictor set in Ham et al. (2019) is orders of magnitude bigger than ours; that is, they consider entire fields (0–360°E and 55°S to 60°N) of the SST and the heat content anomaly as their predictors. In contrast, our model achieves a similar prediction skill with a confined set of just 18 predictors.

When the GDNN and the QRNN ensembles were evaluated on the time period between 1963 and 2017, the estimation of the predictive uncertainty did not lead to a notably better result than the reference method. However, if the models were evaluated on the time period between 1982 and 2017, the GDNN was able to achieve mostly better estimations of the predictive uncertainty than the reference method. Hence, the poor performance of the GDNN ensemble when the full time series is considered might be connected to the lower data quality, which is present for earlier decades. One example of this is the WWV proxy which is used during 1960–1979. Moreover, we suspect that the QRNN has in general a poor performance because it does not make an assumption about the distribution of the predictive uncertainty. Given that there is a relatively low amount of training data available, this makes it especially difficult to predict the correct position of extreme quantiles (see 2.5% and 97.5% quantile in Figures 4b and 4d) because these are defined just by a few data points in the data set.

In summary, this study showed that it is possible to build a skillful statistical model for the long-lead prediction of the ONI using a small predictor set. Moreover, we found indications that a direct estimation of the predictive uncertainty can improve upon simpler methods that rely on the climatological standard deviation of the ONI and the hindcast ACC skill of the model. In future studies, one might further improve the prediction skill of our GDNN by applying transfer learning methods (Pratt et al., 1991), using a skewed distribution for the predictive uncertainty and/or find predictor variables which have a close relationship to the long-term varying predictability of the ENSO.

**Acknowledgments**
This study was conducted within NinoLearn (doi: 10.5281/zenodo.3700434), an open-source research framework for statistical ENSO prediction. The scripts for this study can be found in the subdirectories `research/GDNN` for the GDNN and `research/MQRNN` for the QRNN ensemble models. The code includes all download and preprocessing routines which are necessary to reproduce this study.

## References

Balmaseda, M. A., Davey, M. K., & Anderson, D. L. T. (1995). Decadal and seasonal dependence of ENSO prediction skill. *Journal of Climate*, *8*(11), 2705–2715. https://doi.org/10.1175/1520-0442(1995)008<2705:DASDOE>2.0.CO;2

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. https://doi.org/10.1002/qj.2063

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 200211 is our capability increasing? *Bulletin of the American Meteorological Society*, *93*(5), 631–651. https://doi.org/10.1175/BAMS-D-11-00111.1

Barnston, A. G., Tippett, M. K., Ranganathan, M., & LHeureux, M. L. (2019). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, *53*(12), 7215–7234. https://doi.org/10.1007/s00382-017-3603-3

Barnston, A. G., Tippett, M. K., van den Dool, H. M., & Unger, D. A. (2015). Toward an improved multimodel ENSO prediction. *Journal of Applied Meteorology and Climatology*, *54*(7), 1579–1595. https://doi.org/10.1175/JAMC-D-14-0188.1

Bentzien, S., & Friederichs, P. (2014). Decomposition and graphical portrayal of the quantile score. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1924–1934. https://doi.org/10.1002/qj.2284

Bunge, L., & Clarke, A. J. (2014). On the warm water volume and its changing relationship with ENSO. *Journal of Physical Oceanography*, *44*(5), 1372–1385. https://doi.org/10.1175/JPO-D-13-062.1

Cannon, A. J. (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers & Geosciences*, *37*(9), 1277–1284. https://doi.org/10.1016/j.cageo.2010.07.005

Cannon, A. J. (2018). Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. *Stochastic Environmental Research and Risk Assessment*, *32*(11), 3207–3225. https://doi.org/10.1007/s00477-018-1573-6

Chen, D., & Cane, M. A. (2008). El Niño prediction and predictability. *Journal of Computational Physics*, *227*(7), 3625–3640.

Clarke, A. J. (2014). El Niño physics and El Niño predictability. *Annual Review of Marine Science*, *6*(1), 79–99. https://doi.org/10.1146/annurev-marine-010213-135026

Diaz, H. F., Hoerling, M. P., & Eischeid, J. K. (2001). ENSO variability, teleconnections and climate change. *International Journal of Climatology*, *21*(15), 1845–1862. https://doi.org/10.1002/joc.631

Dijkstra, H. A., Petersik, P., Hernández-García, E., & López, C. (2019). The application of machine learning techniques to improve El Niño prediction skill. *Frontiers of Physics*, *7*, 1–13. https://doi.org/10.3389/fphy.2019.00153

Friederichs, P., & Hense, A. (2007). Statistical downscaling of extreme precipitation events using censored quantile regression. *Monthly Weather Review*, *135*(6), 2365–2378. https://doi.org/10.1175/MWR3403.1

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep Learning for Multiyear ENSO Forecasts Nature. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Hameed, S. N., Jin, D., & Thilakan, V. (2018). A model for super El Niños. *Nature Communications*, *9*(1), 2528. https://doi.org/10.1038/s41467-018-04803-7

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*, *30*(20), 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montegut, C., Behera, S. K., Luo, J.-J., et al. (2010). Influence of the state of the Indian Ocean Dipole on the following years El Niño. *Nature Geoscience*, *3*(3), 168–172. https://doi.org/10.1038/ngeo760

Jin, F.-F. (1997). An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *Journal of the Atmospheric Sciences*, *54*(7), 811–829. https://doi.org/10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-Year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*(448), 1296–1310. https://doi.org/10.1080/01621459.1999.10473882

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In I. Guyon (Ed.), *Advances in neural information processing systems* (Vol. 30, pp. 6402–6413). San Diego, CA: Neural Information Processing Systems Foundation, Inc.

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

Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., & Schellnhuber, H. J. (2014). Very early warning of next El Niño. In *Proceedings of the National Academy of Sciences*, (pp. 201323058). Washington, DC: United States National Academy of Sciences. https://doi.org/10.1073/pnas.1323058111

McDermott, P. L., & Wikle, C. K. (2019). Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data. *Entropy*, *21*(2), 184. https://doi.org/10.3390/e21020184

Morgan, N., & Bourlard, H. (1990). Generalization and parameter estimation in feedforward nets: Some experiments, *Advances in neural information processing systems* (Vol. 2, pp. 630–637). San Diego, CA: Neural Information Processing Systems Foundation, Inc.

Nooteboom, P. D., Feng, Q. Y., Lpez, C., Hernndez-Garca, E., & Dijkstra, H. A. (2018). Using network theory and machine learning to predict El Niño. *Earth System Dynamics*, *9*(3), 969–983. https://doi.org/10.5194/esd-9-969-2018

Pratt, L. Y., Mostow, J., Kamm, C. A., & Kamm, A. A. (1991). Direct transfer of learned information among neural networks, *Aaai* (vol. 91, pp. 584–589). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Prechelt, L. (1998). Early stopping—But when?*Neural networks: Tricks of the trade* (pp. 55–69). Berlin: Springer.

Radebach, A., Donner, R. V., Runge, J., Donges, J. F., & Kurths, J. (2013). Disentangling different types of El Niño episodes by evolving climate network analysis. *Physical Review E*, *88*, 52,807. https://doi.org/10.1103/PhysRevE.88.052807

Rodríguez-Méndez, V., Eguíluz, V. M., Hernández-García, E., & Ramasco, J. J. (2016). Percolation-based precursors of transitions in extended systems. *Scientific Reports*, *6*, 29,552. https://doi.org/10.1038/srep29552

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System version 2. *Journal Of Climate*, *27*(6), 2185–2208.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Suarez, M. J., & Schopf, P. S. (1988). A delayed action oscillator for ENSO. *Journal of the Atmospheric Sciences*, *45*(21), 3283–3287. https://doi.org/10.1175/1520-0469(1988)045<3283:ADAOFE>2.0.CO;2

Tang, Y., Zhang, R.-H., Liu, T., Duan, W., Yang, D., Zheng, F., et al. (2018). Progress in ENSO prediction and predictability study. *National Science Review*, *5*(6), 826–839. https://doi.org/10.1093/nsr/nwy105

Tangang, F. T., Hsieh, W. W., & Tang, B. (1997). Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Climate Dynamics*, *13*(2), 135–147. https://doi.org/10.1007/s003820050156

Tangang, F. T., Hsieh, W. W., & Tang, B. (1998). Forecasting regional sea surface temperatures in the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *Journal of Geophysical Research*, *103*(C4), 7511–7522. https://doi.org/10.1029/97JC03414

Taylor, J. W. (2000). A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, *19*(4), 299–311. https://doi.org/10.1002/1099-131X(200007)19:4<299::AID-FOR775>3.0.CO;2-V

Timmermann, A., An, S.-I., Kug, J.-S., Jin, F.-F., Cai, W., Capotondi, A., et al. (2018). El Niño-Southern Oscillation complexity. *Nature*, *559*(7715), 535–545. https://doi.org/10.1038/s41586-018-0252-6

Wieners, C. E., de Ruijter, W. P. M., Ridderinkhof, W., von der Heydt, A. S., & Dijkstra, H. A. (2016). Coherent tropical Indo-Pacific interannual climate variability. *Journal of Climate*, *29*(11), 4269–4291. https://doi.org/10.1175/JCLI-D-15-0262.1

Wu, A., Hsieh, W. W., & Tang, B. (2006). Neural network forecasts of the tropical Pacific sea surface temperatures. *Neural Networks*, *19*(2), 145–154. https://doi.org/10.1016/j.neunet.2006.01.004