Supporting Information for "Probabilistic forecasting of El Niño using neural network models"

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

Contents of this file

1. Text Sections A-C
2. Section D: Additional Figures S1 to S9

Section A: Network measures

Corresponding author: Paul Petersik, Department of Physics, Utrecht University, Princeton-plein 5, 3584 CC Utrecht, The Netherlands (paul.j.petersik@gmail.com)

[1]Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, The Netherlands

[2]Centre for Complex Systems Studies (CCSS), Department of Physics, Utrecht University, The Netherlands

To determine the ECN, we first compute the Pearson cross-correlation matrix, R, for the SSHA field in the tropical Pacific (30°N-30°S, 120°E-80°W) for each considered time step. Then, the adjacency matrix of the network is defined by:

$$A_{ij} = \Theta(|R_{ij}| - \epsilon) - \delta_{ij} \tag{1}$$

Here, ϵ is a threshold for the correlation coefficient above which two nodes (grid points) are considered to be connected; in this study we choose $\epsilon = 0.9$. With such a high threshold, usually a large part of nodes are not connected to any other node. From the time series of A, we calculate the so-called corrected Hamming distance, $\mathcal{H}^*_{t,t-1}$, between two adjacent time steps as discussed extensively elsewhere (Radebach et al., 2013). This metric can be regarded as the relative change in connectivity of the network from one time step to the following. We assign the computed network metrics to the last month of the time window to make sure not to include any information from the future.

We assume that $\mathcal{H}^*_{t,t-1}$ contains early warning signals for upcoming El Niño/La Niña events due to the increase of spatial correlation during their development. This was recently shown and explained in Rodrguez-Mndez, Eguluz, Hernndez-Garca, and Ramasco (2016). For the understanding of this one needs to apply the view that the ENSO is a dynamical system with a Hopf bifurcation (Jin, 1997) at which oscillations start to grow. Rodrguez-Mndez et al. (2016) indicate that before such a dynamical system passes a bifurcation point, a percolation transition in the the associated evolving complex network takes place. Close to the bifurcation point, the response to perturbations of the system slows down (real part of the Eigenvalues of the Jacobian approach 0). Therefore, temporal correlations in the system start to become stronger. When the system is spatially coupled

the spatial correlation become stronger as well. Hence, the evolving complex climate network becomes more connected. If the system further approaches the bifurcation point a giant component appears which means that most of the nodes are part of the same cluster. The corrected Hamming distance is a variable which displays how strong the network changed in comparison to another network (in our case the network constructed for the 12 month window shifted by one month back in time). In the view of the percolation transition, the Hamming distance shows if and how fast the ongoing transition is.

Section B: Neural network models

In this study, all Multilayer Perceptrons (MLPs) have one hidden layer with 32 neurons. The activation function of the hidden neurons is chosen to be the hyperbolic tangent. The GDNNs and the QRNNs only differ in their output neurons.

The GDNNs have two output neurons, one for the mean, $\hat{\mu}$, and the other for the standard deviation, $\hat{\sigma}$, of a Gaussian distribution (see Fig. S2 for an illustration). The output neuron for the mean $\hat{\mu}_{\theta_i}(\mathbf{x})$ has a linear activation function and the output neuron for the standard deviation $\hat{\sigma}_{\theta_i}(\mathbf{x})$ the softplus function ($f(x) = \ln(1 + e^x)$). To train the members of the GDNN ensemble, the negative log-likelihood (NLL) of a Gaussian distribution is used as loss function. The NLL is computed as follows:

$$L(\hat{\mu}, \hat{\sigma}, y) = -\log P(y|\hat{\mu}, \hat{\sigma}) = \frac{1}{2}\log \hat{\sigma}^2 + \frac{(x - \hat{\mu})^2}{2\hat{\sigma}^2} + constant \tag{2}$$

Here, y is the observation of the predicted variable which is in this study the ONI at some specific lead time. For numerical stability, a very small term of the order $\mathcal{O}(10^{-6})$ was added to $\hat{\sigma}$ during training. The mapping from the predictor set to the predicted mean by the entire ensemble is calculated as follows:

$$\hat{\mu}_*(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \hat{\mu}_{\theta_i}(\mathbf{x}) \tag{3}$$

Here, θ_i refers to the specific parameter setting of member i, i.e. the weights and biases of the ith MLP. The mapping from the predictor set to the predicted standard deviation by the GDNN is computed by

$$\hat{\sigma}^2_*(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} \left(\hat{\sigma}^2_{\theta_i}(\mathbf{x}) - \hat{\mu}^2_{\theta_i}(\mathbf{x})\right) - \hat{\mu}^2_*(\mathbf{x}). \tag{4}$$

The QRNNs have five output neurons that represent the values of the 2.5-%, 16-%, 50-%, 84-% and 97.5-%-quantiles. The output neuron for the median has a linear activation function. All other output neurons have a ReLU as activation such that they can represent a non-negative offset to an adjacent quantile. In case the considered quantile is greater than the median the offset is added to the predicted value of the lower quantile. Hence, the final prediction is generated in a recursive manner. First, the median is predicted and then the predicted offsets for the two adjacent quantiles, i.e. 16% and 84%-quantiles, are subtracted/added from the median, respectively. Afterwards the values for the 2.5%- and 97.5% quantiles can be computed in the same way. However, this time the offsets are subtracted/added from the value of the 16% and 84%-quantiles, respectively.

As loss function the average tilted loss is used. The tilted loss for the qth quantile is defined by

$$L(\hat{y}, y, q) = \max(\{q(y - \hat{y}), (1 - q)(y - \hat{y})\}) \tag{5}$$

where $\hat{y}$ is the value of the predicted quantile and y, again, the value of the observed ONI. The average tilted loss using the quantiles $q \in \{0.025, 0.16, 0.5, 0.84, 0.975\}$ is hence

$$\overline{L(\hat{y}, y)} = \frac{1}{5} \sum_{i=1}^{5} L(\hat{y}, y, q_i). \tag{6}$$

In the ensemble prediction, the output values for each quantile are averaged over all ensembles.

In summary, all networks are trained with the following (hyperparameter) settings:

- hidden layers: 1
- hidden layer neurons: 32
- dropout rate hidden neurons (only training): 0.05

- activation function for hidden layers: hyperbolic tangent
- learning rate: 0.01
- mini-batch size: 100
- patience before early stopping is applied: 25 epochs
- Optimizer: Adam (Kingma & Ba, 2014)

Section C: Ensemble Method

To train and evaluate an ensemble, the entire time series needs to be split into training, validation and test data sets. The data splitting in this study works as follows: First, the entire time series is split into six test periods, namely 1963-1971, 1972-1981, 1982-1991, 1992-2001, 2002-2011 and 2011-2017. Hence, in total, six ensembles (for the six test periods) are needed for each lead time to perform an evaluation of the ensemble for the entire time series. Each of these periods will once serve as test data set. After one of the six periods is split apart, the remaining time series is further divided into five segments. Four segments now serve as the training data for three members and the remaining segment as validation data set. This is done until each segment was once the validation data set. Hence, an ensemble in this study has $3 \times 5 = 15$ members. Fig. S3 illustrates the described data splitting.

Section D: Additional Figures

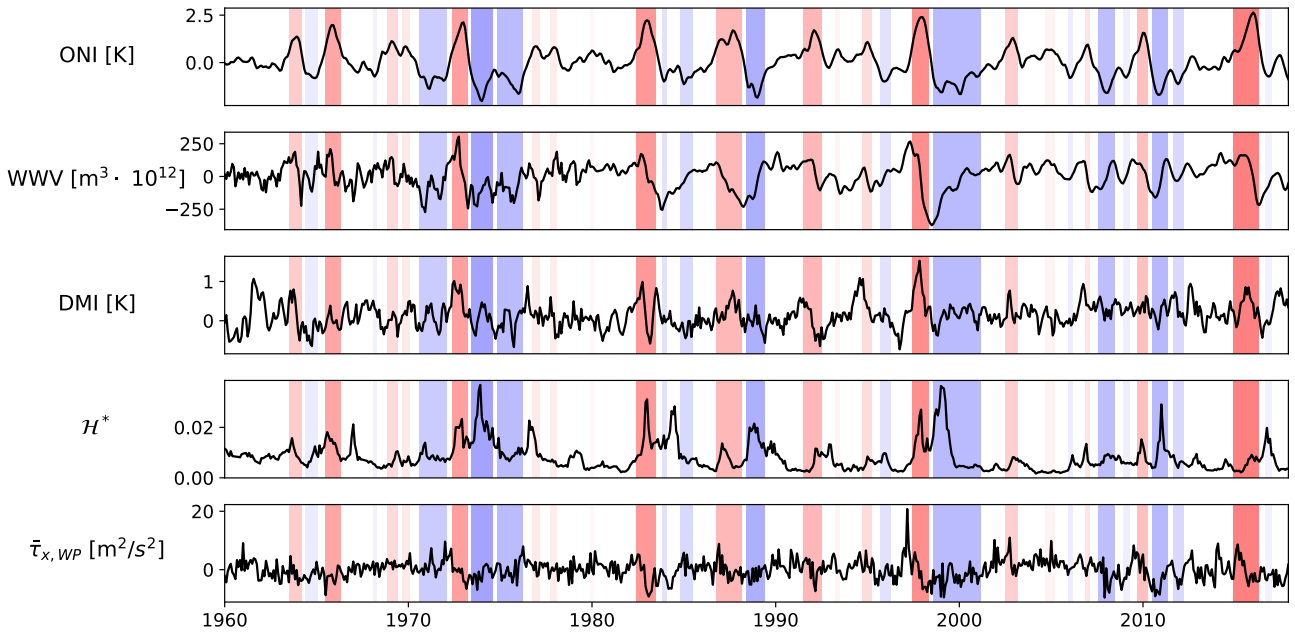


Figure S1. Time series of all predictor variables besides SC. The red background colour indicates time periods during which the ONI was greater than 0.5 K and the blue during which it was smaller than -0.5 K.

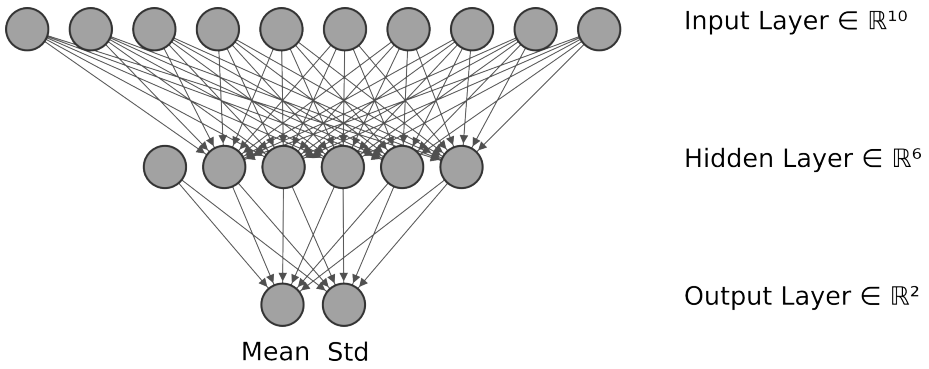


Figure S2. Exemplaric architecture of a GDNN member having ten neurons in the input, six neurons in the hidden and two neurons for the mean and the standard deviation in the output layer. One neuron in the input and the hidden layer has a constant output of 1 to account for the bias term in the subsequent layer.

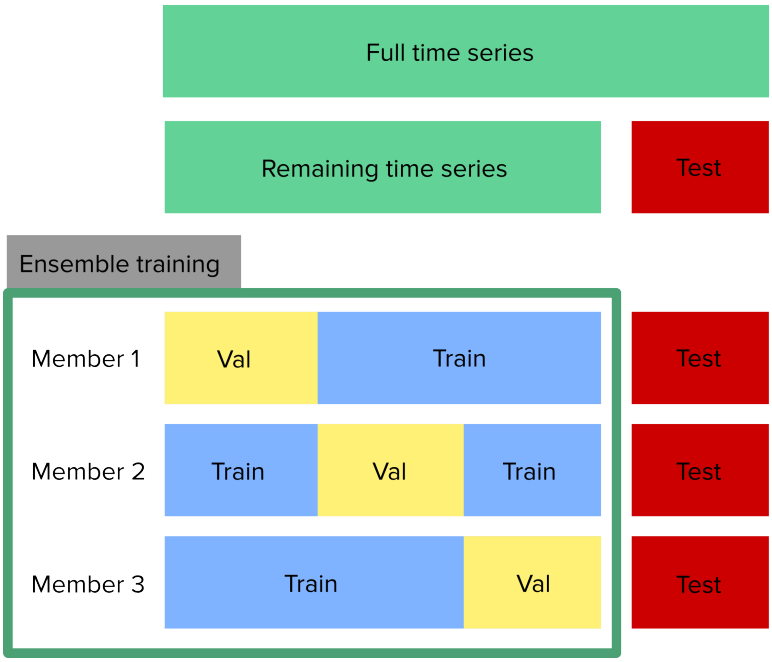


Figure S3. A sketch of the data splitting and ensemble training which is applied in this study. A test data (red) is first hold out for later evaluation. Then, the remaining series is divided (here) into three segements. Each segment is one time the validation data set (yellow). Here, one ensemble member is trained on the two remaining segments (blue).

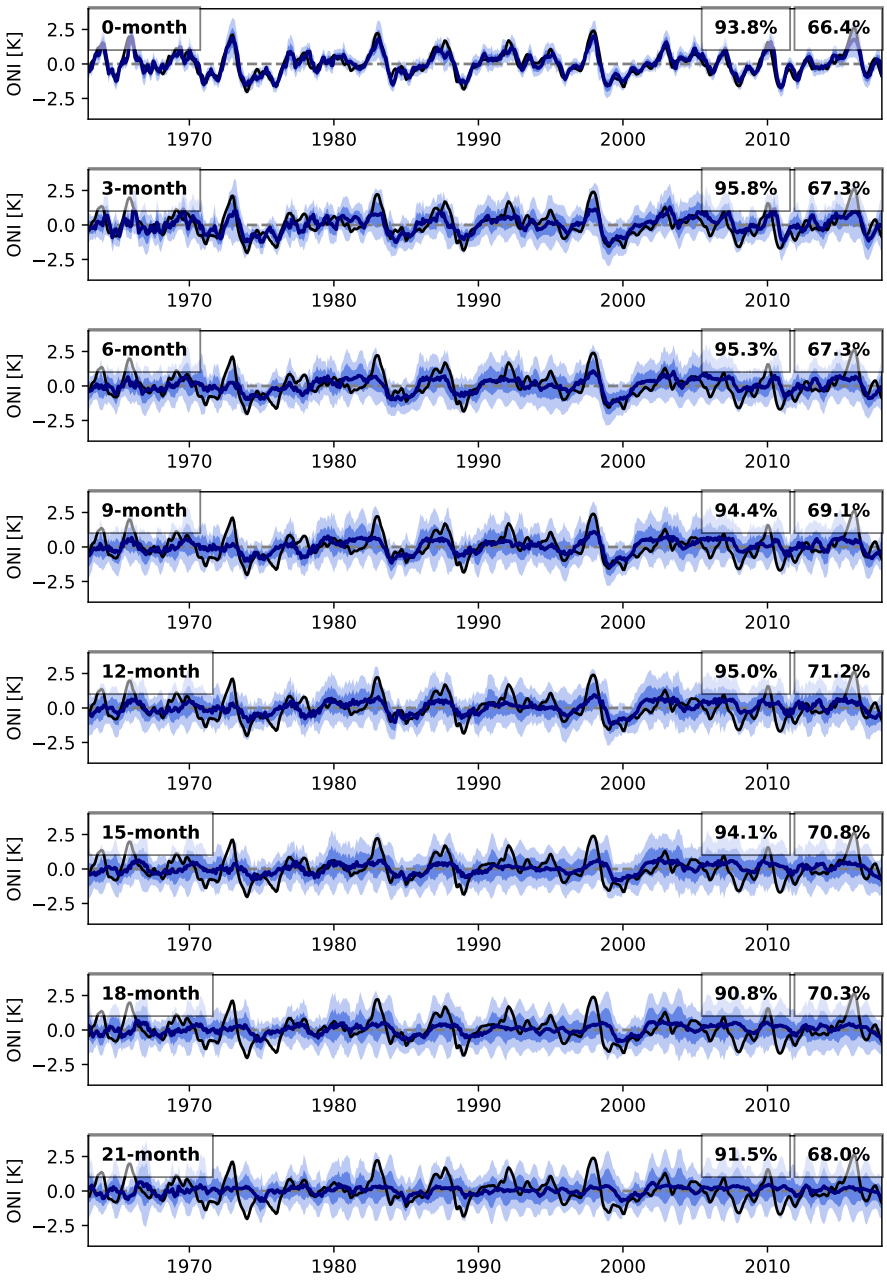


Figure S4. Hindcasts for the GDNN ensemble for all considered lead times. The solid blue line indicates the predicted mean. The darker blue shading displays the area between the 16-% and the 84-% quantiles. The lower lighter shadings indicates area between the 2.5-% and the 16-% quantiles and the upper 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 number) are shown in the upper right corners.

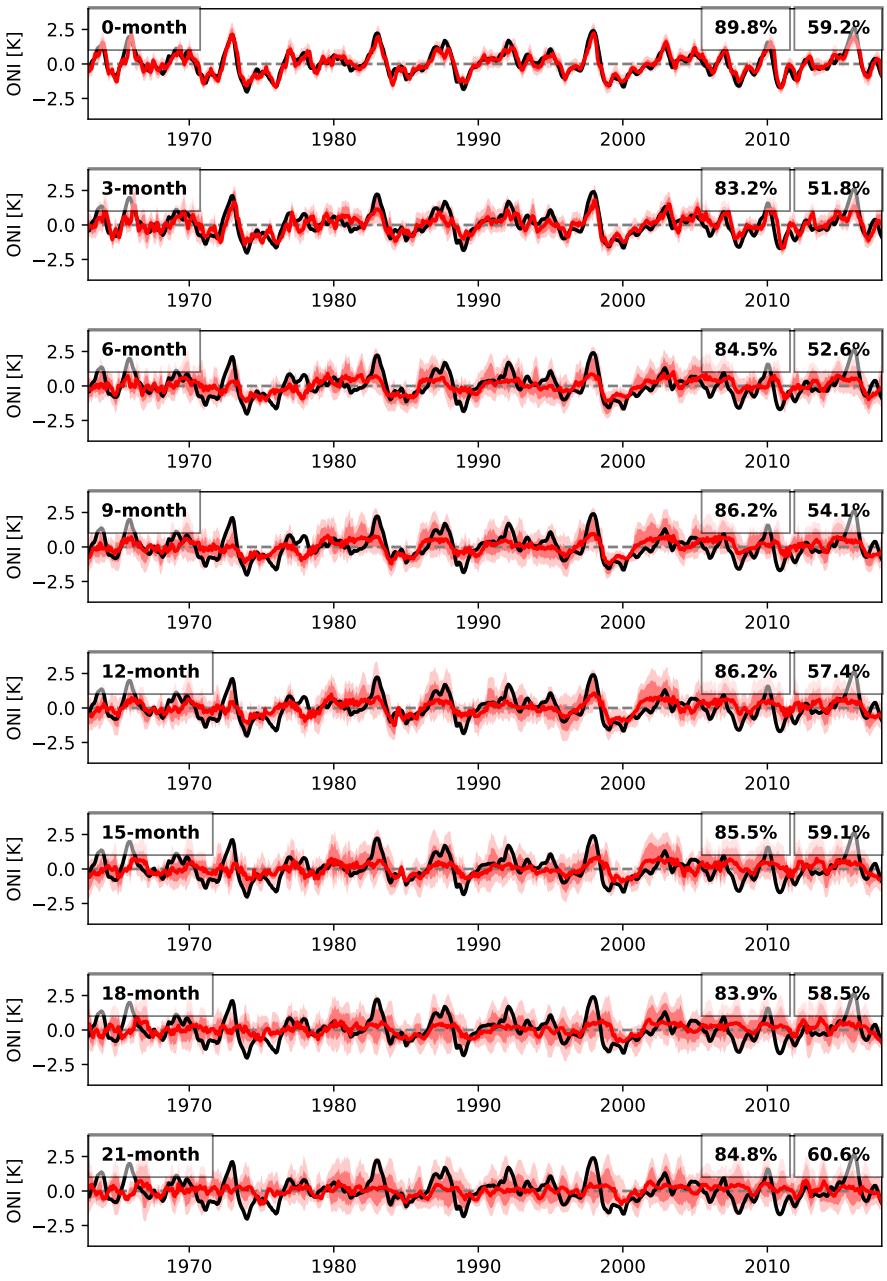
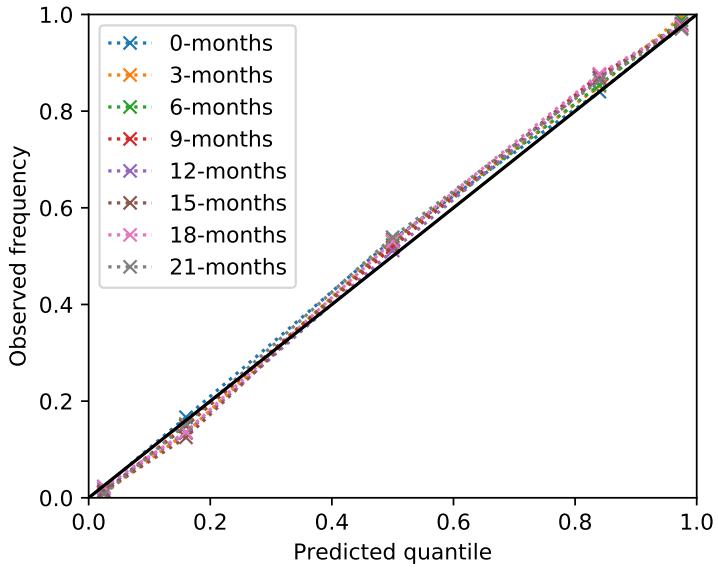
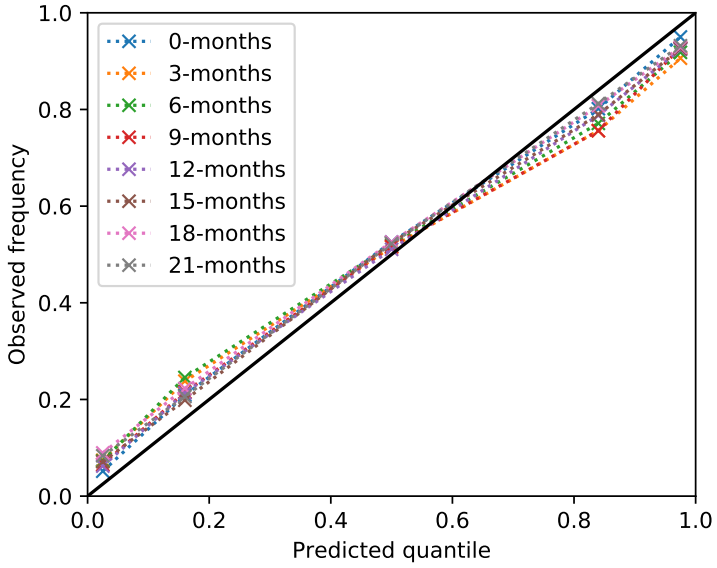


Figure S5. Hindcasts for the QRNN ensemble for all considered lead times. The solid red line indicates the predicted median (50-% quantile). The darker red shading displays the area between the 16-% and the 84-% quantiles. The lower lighter shadings indicates area between the 2.5-% and the 16-% quantiles 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 number) are shown in the upper right corners.

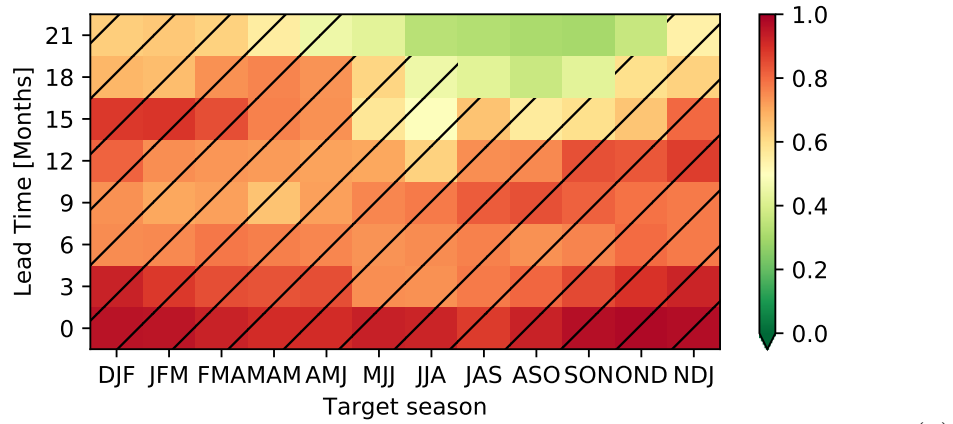


(a)

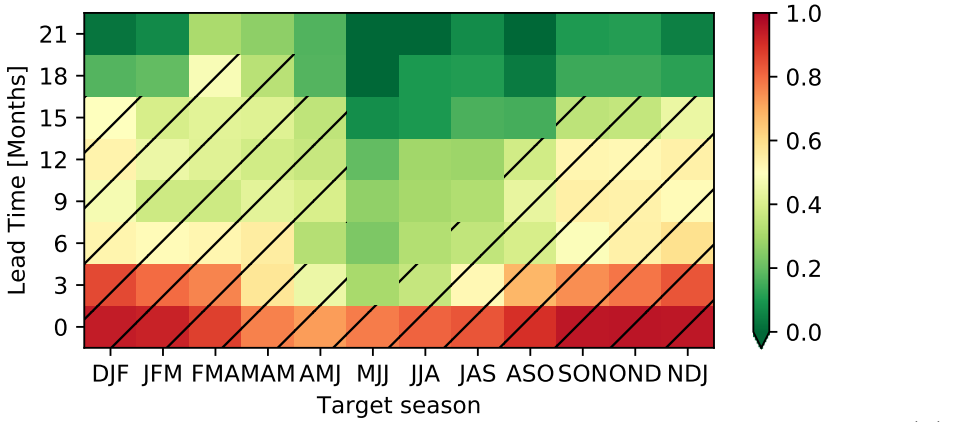


(b)

Figure S6. Reliability diagram for the GDNN (a) and the QRNN (b) ensemble for all considered lead times. The black line indicates perfect reliability.



(a)



(b)

Figure S7. Seasonal ACC-skill of the QRNN (a) during 1982-2001 (El Niño-like period) and (b) durng 1963-1981 and 2001-2017 (La Niña-like period). Hatches represent an ACC skill that is significant with a 95%-confidence.

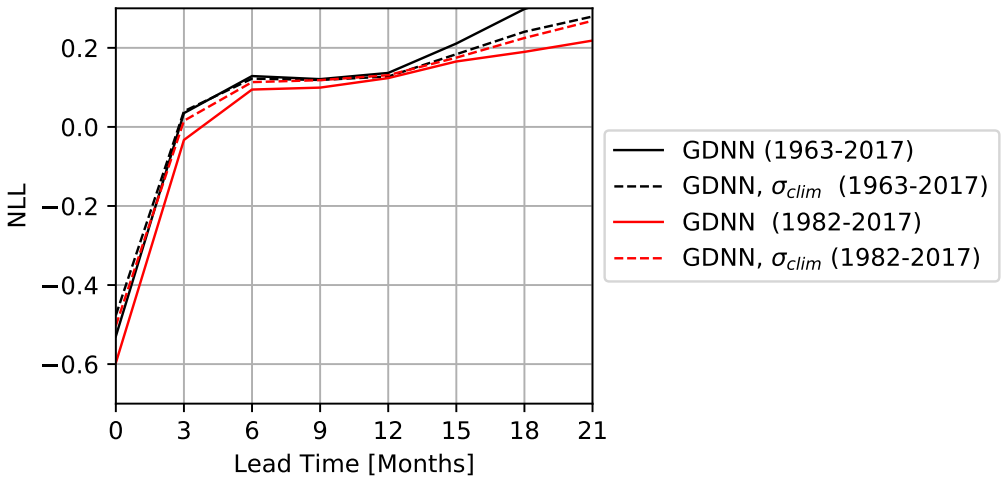


Figure S8. NLL-skill for various lead times for the GDNN ensembles for the period between 1963-2017 (black) and 1982-2017 (red) using the predicted standard deviation and the standard deviation which is estimated by Eq. (5) of (Barnston et al., 2015) (dashed).

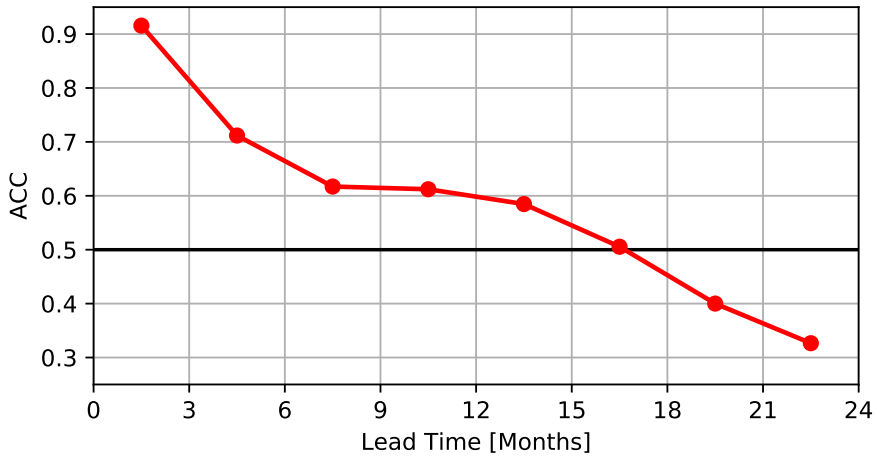


Figure S9. ACC-skill of the GDNN ensembles for the period between 1984-2017. Lead times are in this plot defined as in Ham et al. (2019). This plot hence is directly comparable to Fig. 2a of Ham et al. (2019).

References

Barnston, A. G., Tippett, M. K., van den Dool, H. M., & Unger, D. A. (2015, April). Toward an Improved Multimodel ENSO Prediction. *J. Appl. Meteor. Climatol.*, *54*(7), 1579–1595. Retrieved 2020-01-05, from https://journals.ametsoc.org/doi/full/10.1175/JAMC-D-14-0188.1 doi: 10.1175/JAMC-D-14-0188.1

Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019, September). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. Retrieved 2019-11-11, from https://www.nature.com/articles/s41586-019-1559-7 doi: 10.1038/s41586-019-1559-7

Jin, F.-F. (1997, April). An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model. *J. Atmos. Sci.*, *54*(7), 811–829. Retrieved 2019-04-08, from https://journals.ametsoc.org/doi/full/10.1175/1520-0469%281997%29054%3C0811%3AAEORPF%3E2.0.CO%3B2 doi: 10.1175/1520-0469(1997)054⟨0811:AEORPF⟩2.0.CO;2

Kingma, D. P., & Ba, J. (2014, December). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. Retrieved 2019-08-07, from http://arxiv.org/abs/1412.6980 (arXiv: 1412.6980)

Radebach, A., Donner, R. V., Runge, J., Donges, J. F., & Kurths, J. (2013, November). Disentangling different types of El Nio episodes by evolving climate network analysis. *Physical Review E*, *88*(5). Retrieved 2018-09-19, from https://link.aps.org/doi/10.1103/PhysRevE.88.052807 doi: 10.1103/PhysRevE.88.052807

Rodrguez-Mndez, V., Eguluz, V. M., Hernndez-Garca, E., & Ramasco, J. J. (2016, July). Percolation-based precursors of transitions in extended systems. *Scientific Reports*, *6*, 29552.

Retrieved 2019-03-14, from https://www.nature.com/articles/srep29552 doi: 10.1038/srep29552