Supporting Information for "Deep Residual Convolutional Neural Network Combining Dropout and Transfer Learning for ENSO Forecasting"

Jie Hu[1,2,3] *, Bin Weng[1,2,3] †, Tianqiang Huang[1,2,3], Jianyun Gao[4], Feng Ye[1,2,3], Lijun You[5]

[1]College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350007, China

[2]Digital Fujian Institute of Big Data Security Technology, Fuzhou 350007, China

[3]Engineering Technology Research Center for Public Service Big Data Mining and Application of Fujian Province, Fuzhou 350007, China

[4]Fujian Key Laboratory of Severe Weather, Fujian Institute of Meteorological Sciences, Fuzhou 350007, China

[5]Fujian Meteorological Information Center, Fuzhou 350007, China

1. Text S1
2. Text S2
3. Text S3

Corresponding author: Tianqiang Huang, fjhtq@fjnu.edu.cn

Corresponding author: Jianyun Gao, fzgaojyun@163.com

*Co-first authors

†Co-first authors

4. Figures S1 to S9
5. Table S1 to S3

Text S1. Dropout and Transfer Learning

The dropout technique is used to alleviate the overfitting issue. In this work we use the base version (Srivastava et al., 2014), i.e., in each iteration, each node may be dropped out with probability p according to the Bernoulli distribution. The p is the probability that the node is set to zero, it's then scaled by a factor of $1/(1-p)$. This method can facilitate the interaction between feature detectors (hidden layer nodes). Detector interaction means that some detectors rely on other detectors to function. In other words, it lets the activation value of a neuron stop working with a certain probability as forwarding propagation. This can make the model more general because it does not rely too much on some local characteristics.

For transfer learning (Zhuang et al., 2020), the feature space of the data is represented by attributes ($X_s = X_t$) and labels (Y_s, Y_t), with s representing the source object and t representing the target object. Y_s and Y_t are equal for homogeneous transfer learning and unequal for heterogeneous transfer learning. The most straightforward method is applying the source domain model directly to the target domain before training. For example, the ONI (type) is predicted by using homogeneous (heterogeneous) transfer learning, which uses the model parameters of predicting the Niño3.4 index as the initialization parameters before training.

Text S2. Indexes forecast

The batch size is 400 (20) and epoch is 100 (50) during transfer learning (training). The learning rate is set to 0.001 and a learning rate scheduler is used. To compare the prediction performance of Res-CNN, the North American Multi-model Ensemble phase 1 (Kirtman et al., 2014), Scale Interaction Experiment-Frontier (SINTEX-F) (Luo et al., 2008), and CNN are used during the valid period from 1984 to 2017. To predict ONI, we divide the valid data into four-time slices, 1982-1991, 1992-2001, 2002-2011, and 2012-2017. First, one slice data is randomly taken as a test set, then the first 20 years of the remaining slice data are considered as a training set, while the remaining time slice data as a validation set. Since the test sets are independent, there is no overlap between training data, validation data, and test data. The batch size is 100, the epoch is 100, and the learning rate is 0.0005.

Text S3. Types forecast

We measure the performance using accuracy, which is the number of correct predictions divided by the total number of predictions. All types are defined by the Niño3 index (SSTA averaged over 150°- 90°W, 5°S-5°N), denoted as N_3, and the Niño4 index (SSTA averaged over 160°E-150°W, 5°S-5°N), denoted as N_4. The calculation of classification is as follows:

$$
\begin{aligned}
r &= \sqrt{(N_3+N_4)^2+(N_3-N_4)^2} = \sqrt{2\left(N_3^2+N_4^2\right)} \\
\theta &= \begin{cases} \arctan\frac{(N_3-N_4)}{(N_3+N_4)} & N_3+N_4>0 \\ \arctan\frac{(N_3-N_4)}{(N_3+N_4)}-180 & N_3+N_4<0 \end{cases} \\
\theta &\in \begin{cases} (15^\circ, 90^\circ) & \text{EP EI Niño} \\ (-15^\circ, 15^\circ) & \text{MIX EI Niño} \\ (-90^\circ, -15^\circ) & \text{CP EI Niño} \\ (-165^\circ, -90^\circ) & \text{EP La Niña} \\ (-195^\circ, -165^\circ) & \text{MIX La Niña} \\ (-270^\circ, -195^\circ) & \text{CP La Niña} \\ \text{other} & \text{NY} \end{cases}
\end{aligned} \tag{1}
$$

Finally, the ENSO events occur when r in the December-January-February (DJF) season is greater than their standard deviation (Ham et al., 2019).

References

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year enso forecasts [Journal Article]. *Nature*, *573*(7775), 568-572. Retrieved from `https://www.ncbi.nlm.nih.gov/pubmed/31534218` doi: 10.1038/s41586-019-1559-7

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., ... others (2014). The north american multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. doi: 10.1175/bams-d-12-00050.1

Luo, J.-J., Masson, S., Behera, S. K., & Yamagata, T. (2008). Extended enso predictions using a fully coupled ocean–atmosphere model [Journal Article]. *Journal of Climate*, *21*(1), 84-93. doi: 10.1175/2007jcli1412.1

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.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), 43–76. doi: 10.1109/jproc.2020.3004555

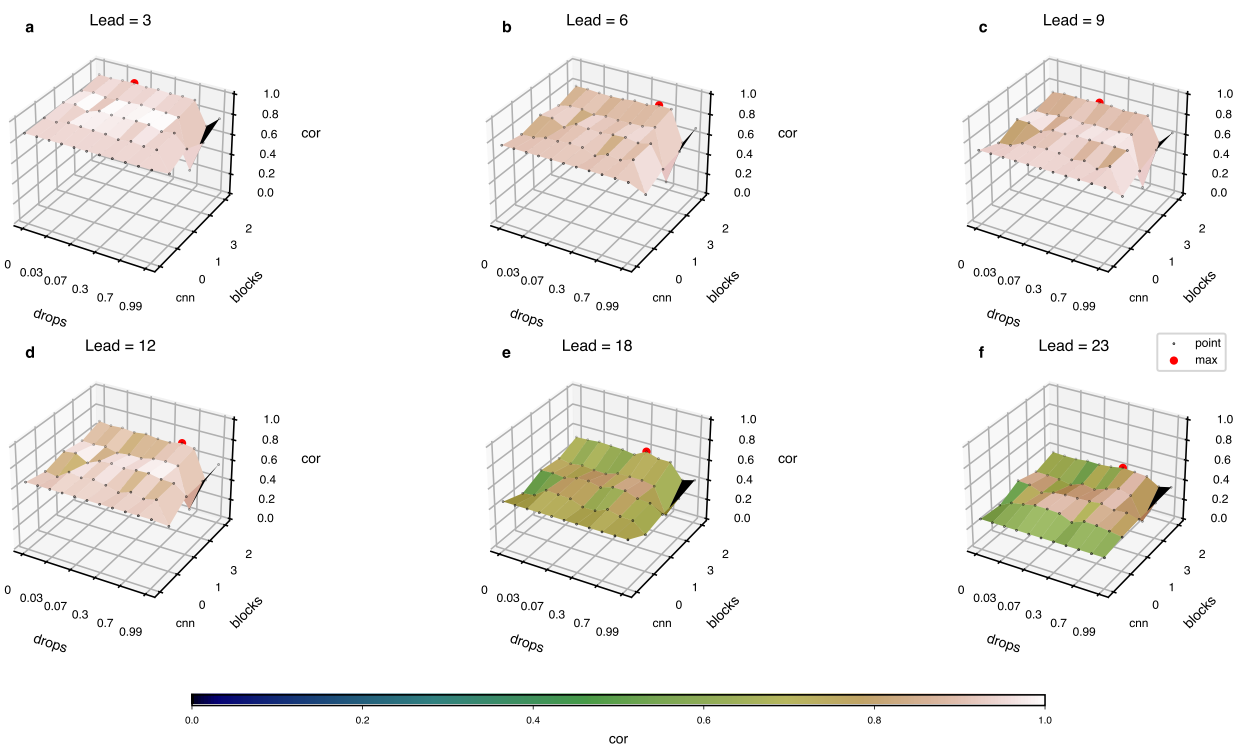

Figure S1. Model selection. The optimal model for a given lead month is selected from many different dropout rates and the number of residuals connected. Here only the advance months are shown as 3, 6, 9, 12, 18, 23, and the corresponding is a-f. The best number of residual blocks is 2, and the best dropout rate is between 0.05 and 0.5.

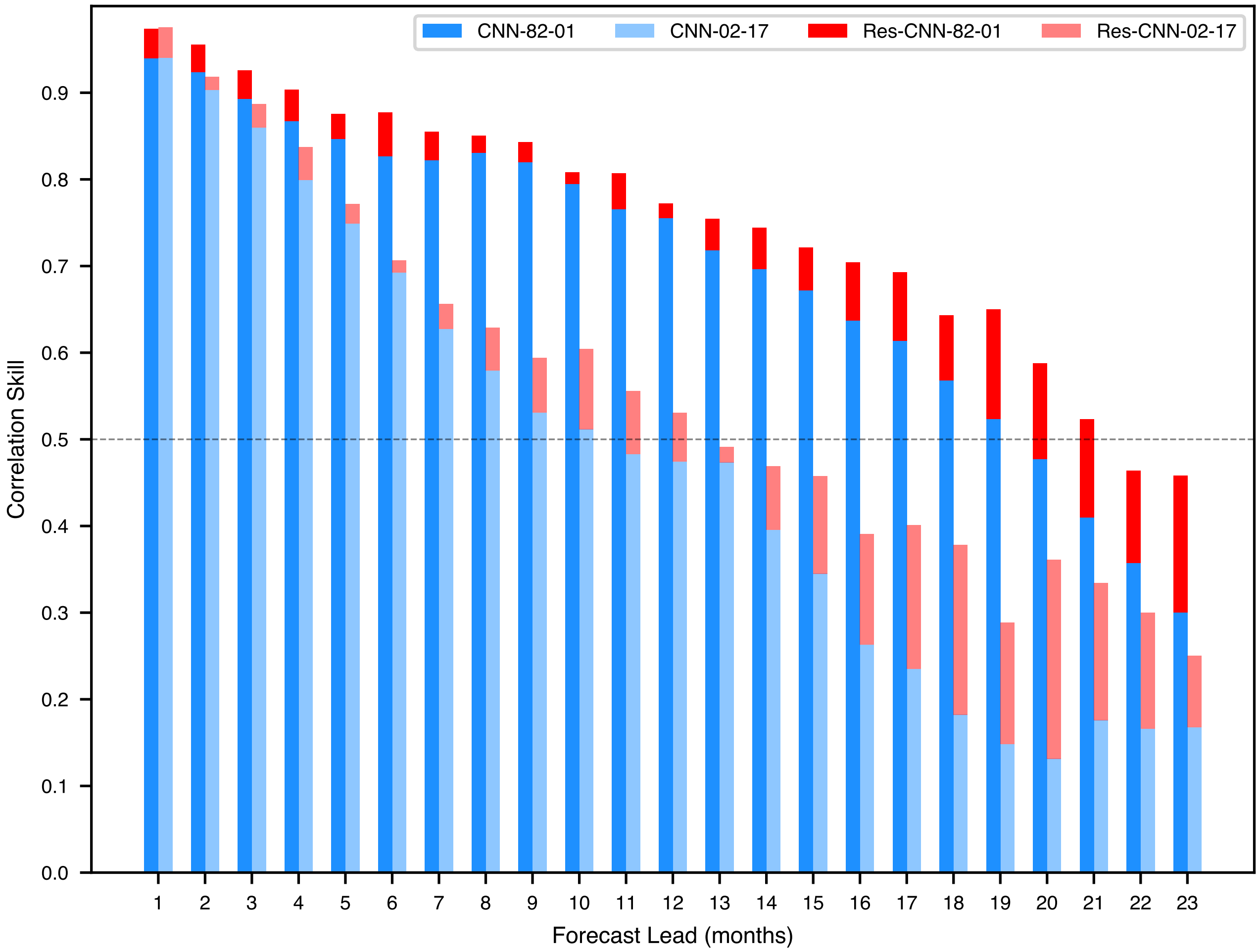

Figure S2. Correlation skill of the CNN and the Res-CNN for various lead times and decades. The all-season correlation skill of the three-month-moving-averaged Niño3.4 index from 1982 to 2001 (deep) and from 2002 to 2017 (shallow) as a function of the prediction lead month using the CNN model (dodger blue) and the Res-CNN model (red). The black dashed line indicates that the correlation coefficient is equal to 0.5.

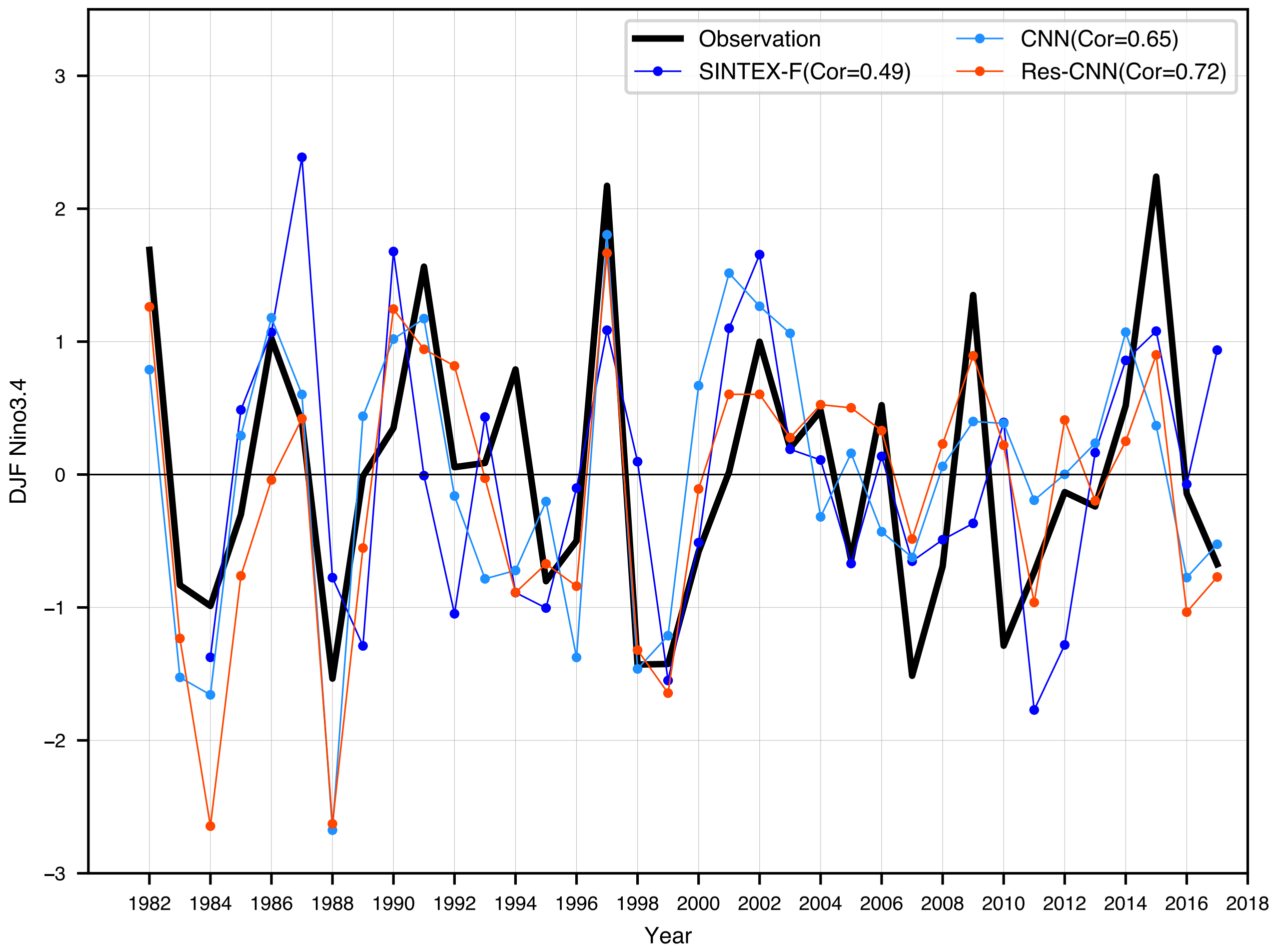

Figure S3. Time series 18 months in advance. Time series of DJF season Niño3.4 indexes for 18-month-lead prediction using the CNN model (dodger blue) and the Res-CNN model(red). Cor means correlation skill for DJF season from 1982 to 2017.

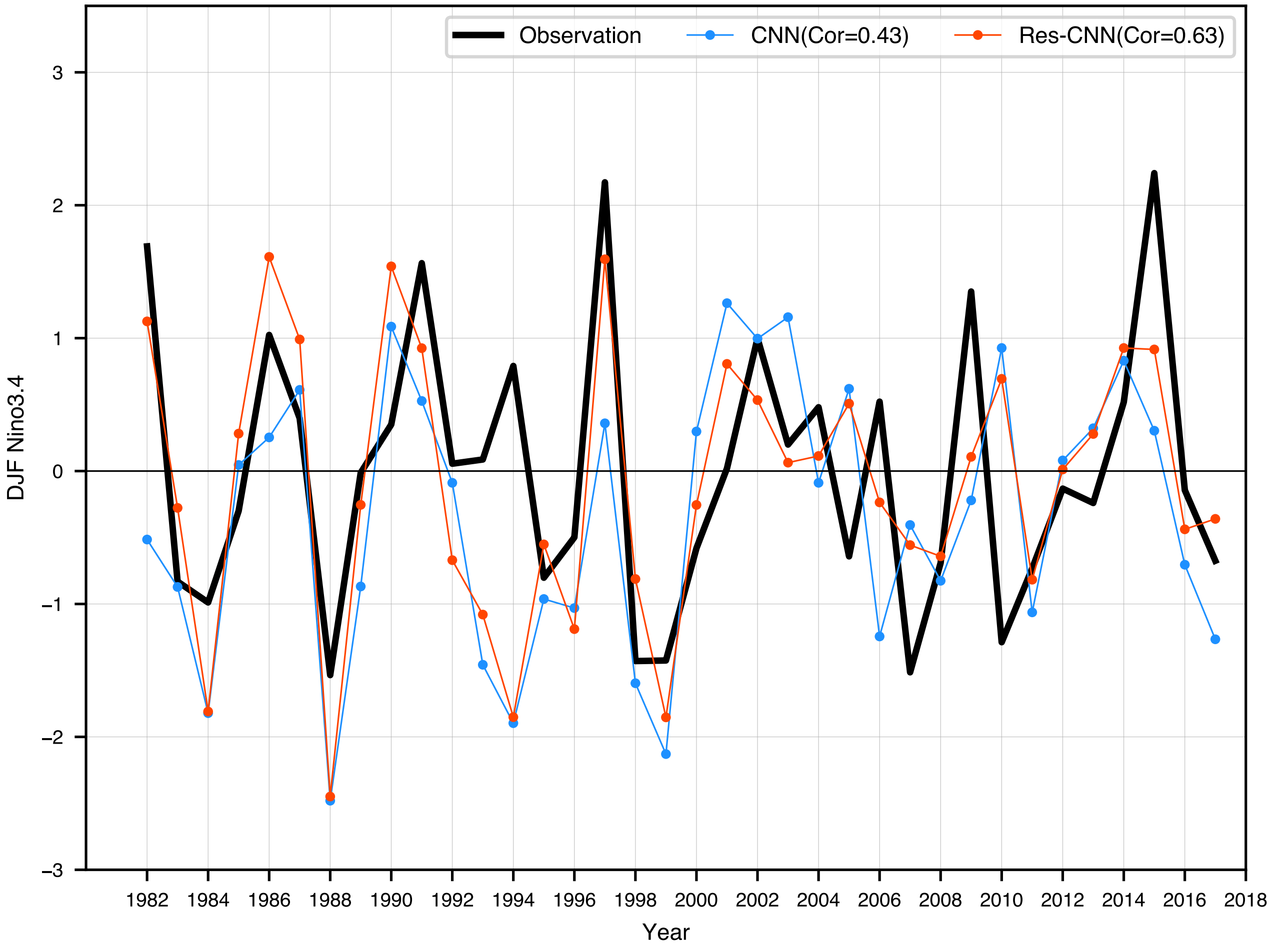

Figure S4. Niño3.4 prediction 20-month in advance. Time-series of DJF season Niño3.4 indexes for 20-month-lead prediction using the CNN model (dodger blue), and the Res-CNN model(red). Cor means correlation skill for DJF season from 1982 to 2017.

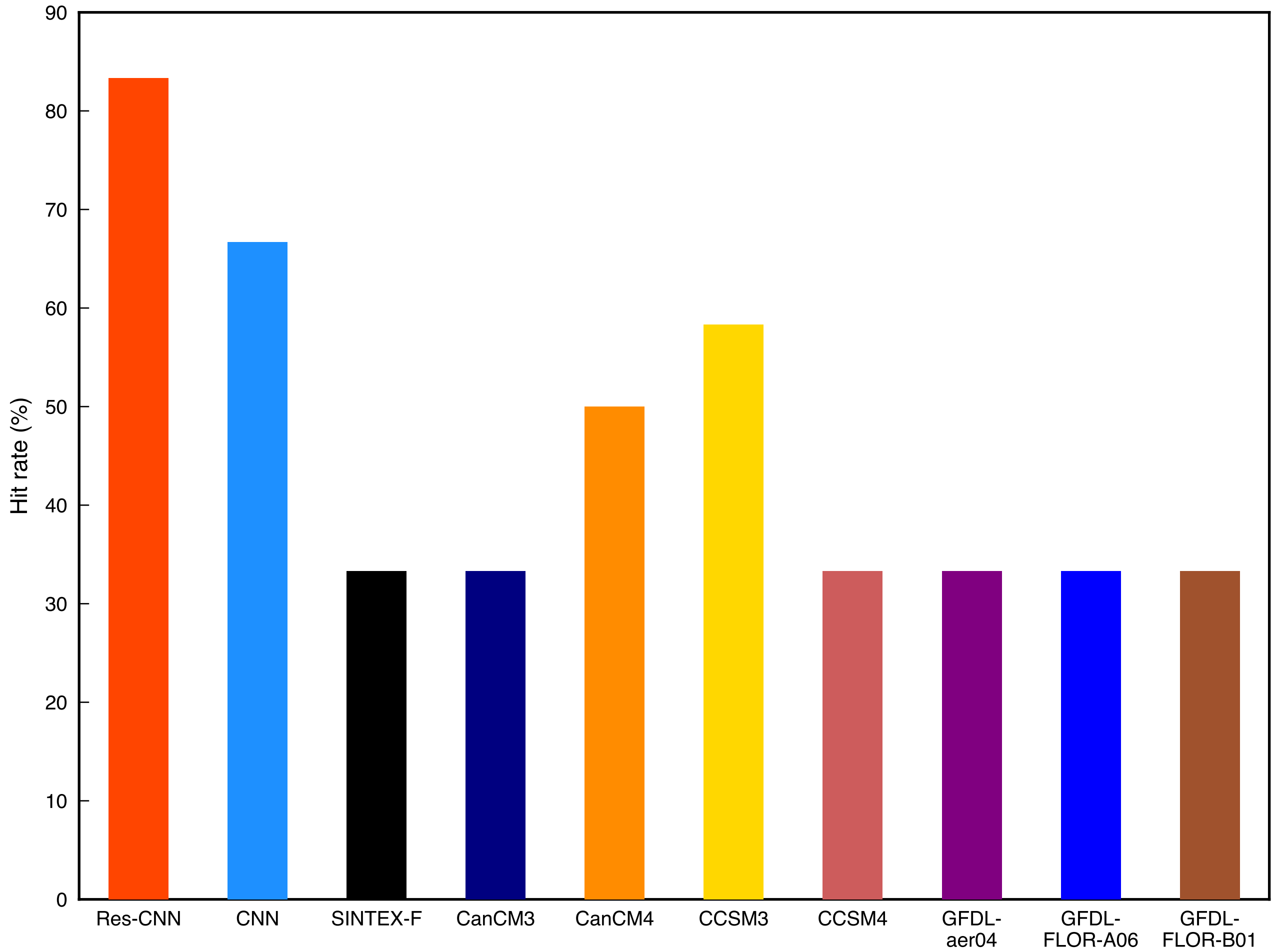

Figure S5. Prediction of EI Niño type. The prediction accuracy of EI Niño types (EP, CP, MIX) 12 months in advance using Res-CNN (red), CNN (dodger blue), and other models from 1982 to 2017.

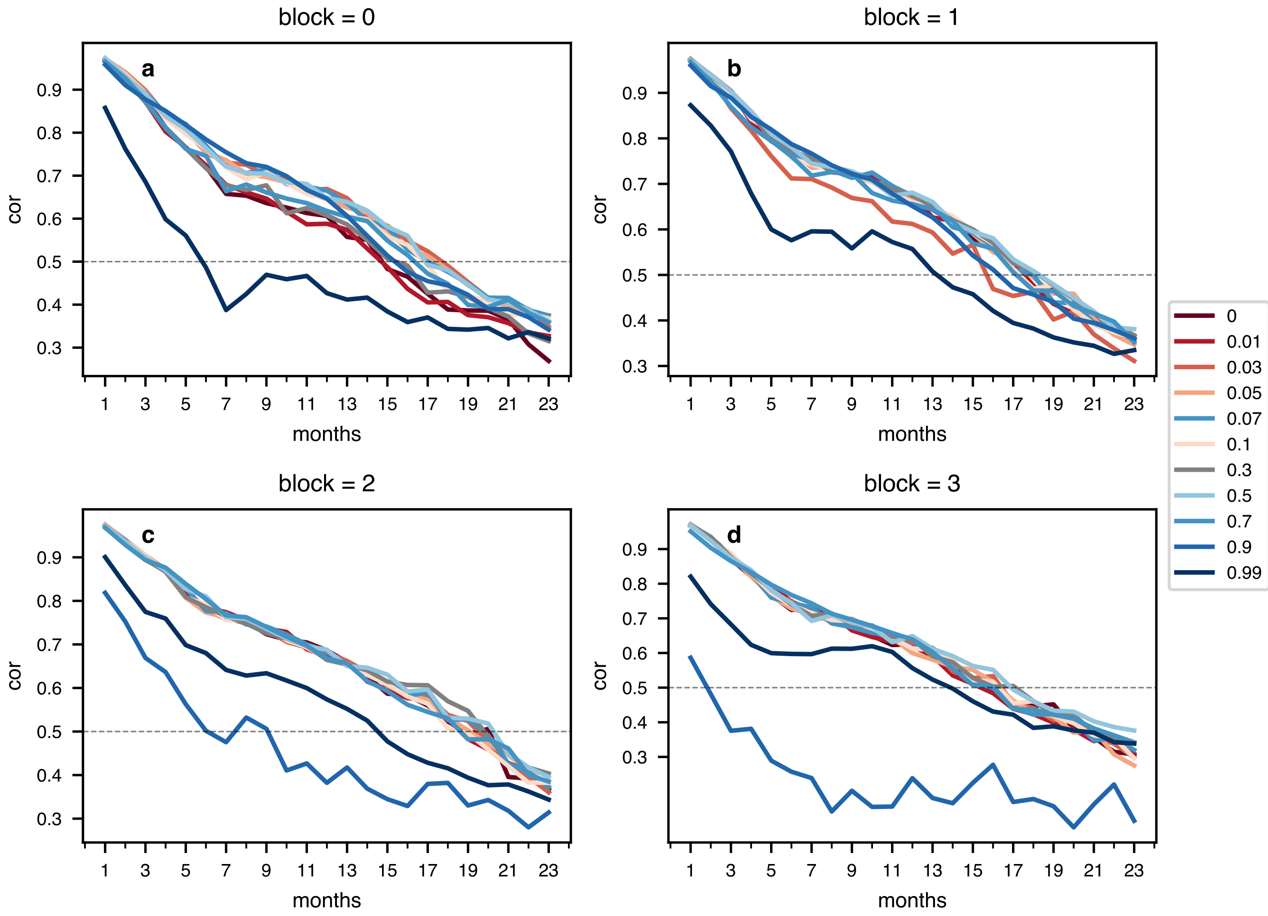

Figure S6. The impact of the dropout. The effect of the dropout rate under different skip connection conditions. Here the skip connections of 0-3 are shown as a-d, respectively. The black dashed line indicates that the correlation coefficient is equal to 0.5. Notice that the dropout rate of 0.99 is better than the dropout rate of 0.9 for c and d, probably because the number of skip connections is higher, and its effect is greater than the dropout.

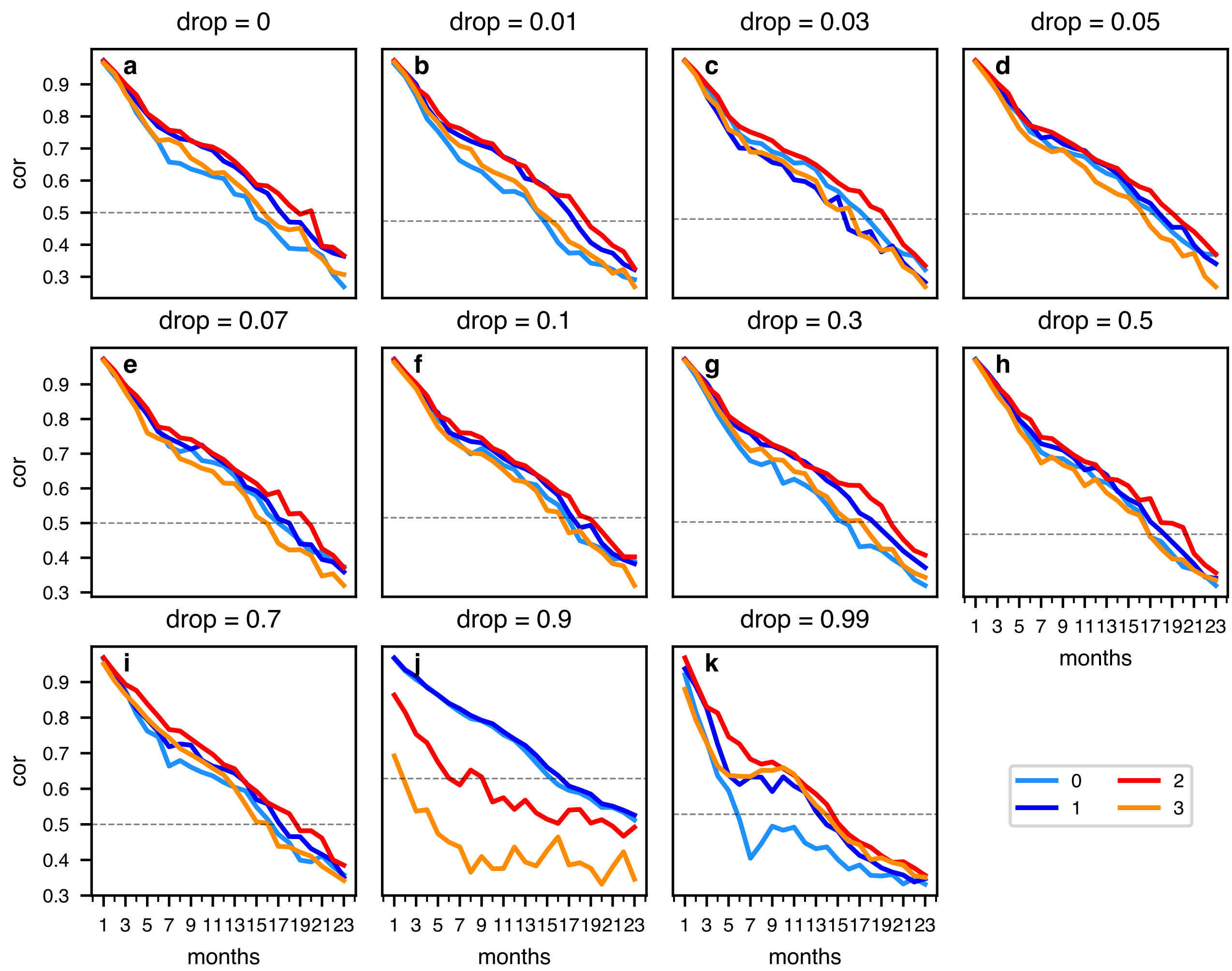

Figure S7. The impact of skip connection. The effect of the skip connection under different dropout rate conditions. Here the dropout rates are 0, 0.01, 0.03, 0.05, 0.07, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99, shown in a-k, respectively. The black dashed line indicates that the correlation coefficient is equal to 0.5.

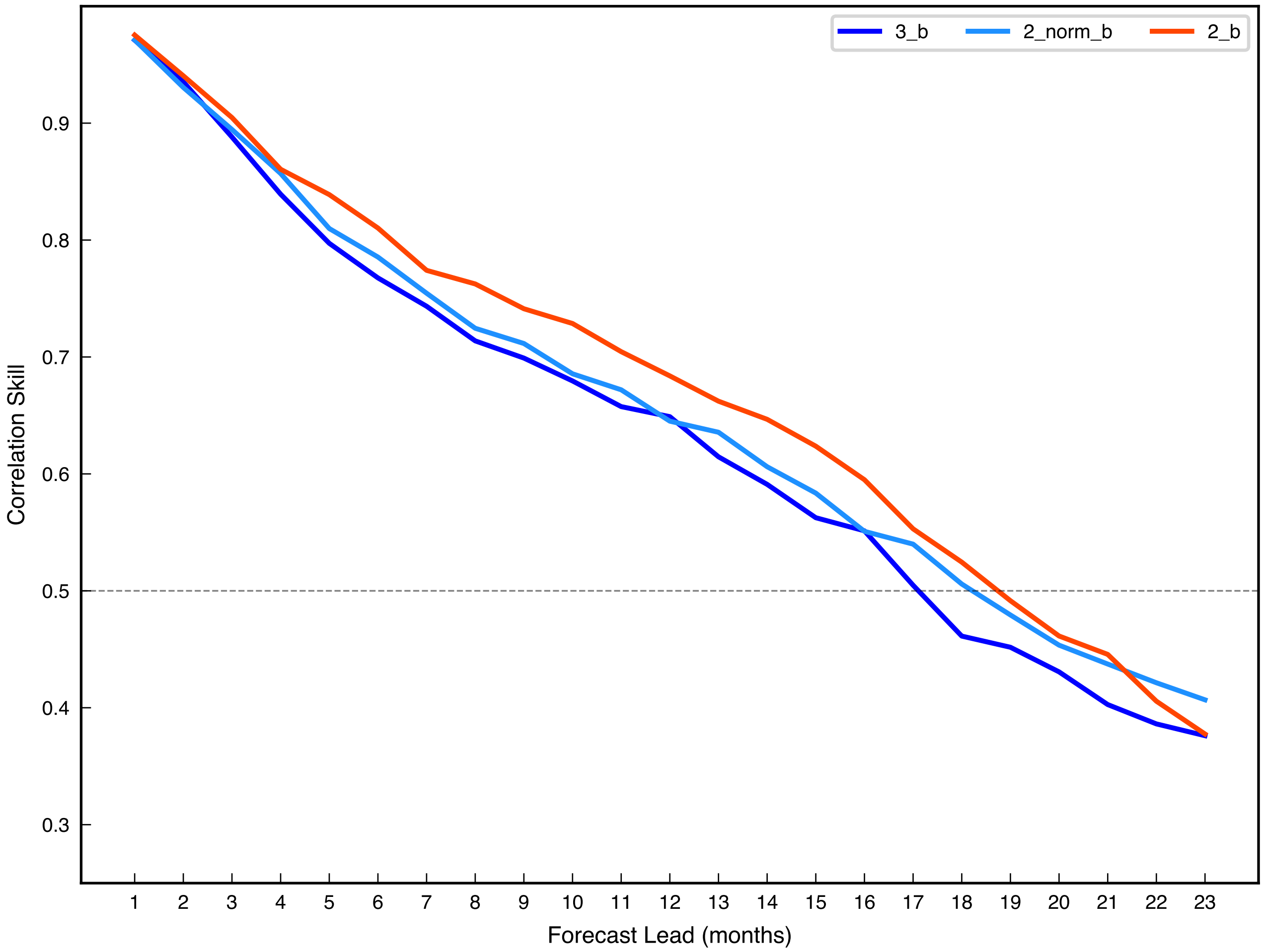

Figure S8. Compares the skill between un-normalized 3 residual connections (blue), un-normalized 2 residual connections (red), and normalized 2 residual connections (dodger blue). The black dashed line indicates that the correlation coefficient is equal to 0.5.

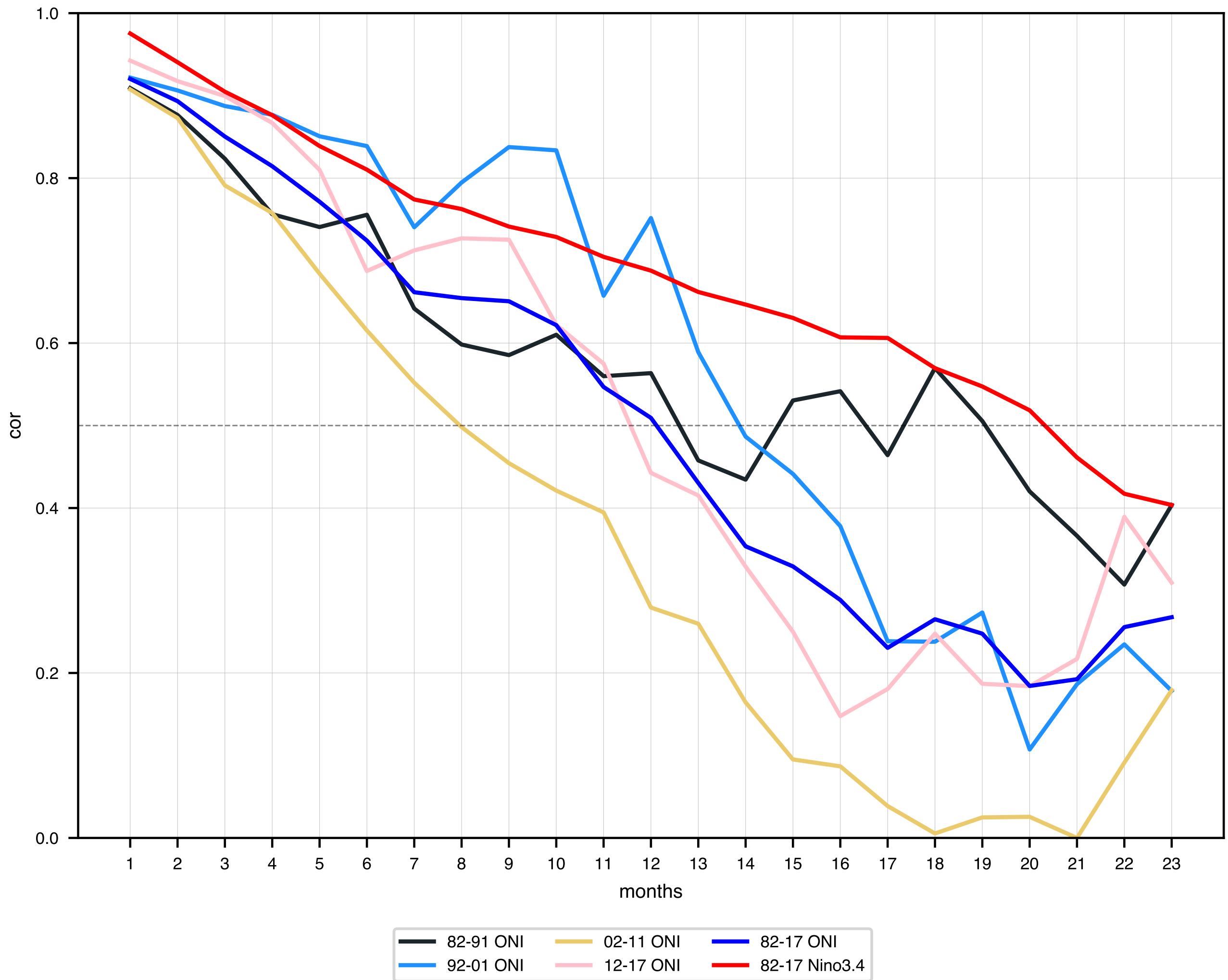

Figure S9. All-season correlation skill of the ONI and Niño3.4 index. The all-season correlation skill of the three-month-moving-averaged ONI from 1982 to 1991(black), ONI from 1992 to 2001(dodger blue), ONI from 2002 to 2011(khaki), ONI from 2012 to 2017(pink), ONI from 1982 to 2017(blue), Niño3.4 from 1982 to 2017(red) using the Res-CNN without transfer learning for various lead times. The black dashed line indicates that the correlation coefficient is equal to 0.5.

Table S1. Correlation skill - Res-CNN

	Target season											
Forecast lead	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF
1	0.98	0.97	0.99	0.96	0.95	0.97	0.97	0.98	0.97	0.97	0.99	0.99
2	0.95	0.96	0.93	0.89	0.89	0.90	0.95	0.94	0.97	0.97	0.97	0.97
3	0.94	0.92	0.88	0.84	0.85	0.83	0.88	0.92	0.93	0.97	0.97	0.95
4	0.93	0.90	0.86	0.80	0.78	0.82	0.83	0.87	0.91	0.92	0.95	0.93
5	0.92	0.88	0.82	0.79	0.76	0.80	0.75	0.79	0.86	0.90	0.91	0.90
6	0.89	0.87	0.82	0.71	0.69	0.79	0.77	0.74	0.79	0.83	0.89	0.93
7	0.89	0.87	0.81	0.68	0.64	0.73	0.72	0.77	0.69	0.75	0.83	0.89
8	0.88	0.86	0.83	0.74	0.69	0.71	0.74	0.68	0.74	0.70	0.77	0.80
9	0.85	0.85	0.84	0.70	0.59	0.70	0.72	0.69	0.76	0.72	0.70	0.78
10	0.81	0.81	0.79	0.68	0.66	0.66	0.71	0.72	0.74	0.71	0.74	0.69
11	0.79	0.81	0.82	0.59	0.52	0.62	0.69	0.70	0.74	0.72	0.71	0.74
12	0.80	0.79	0.84	0.64	0.55	0.56	0.63	0.62	0.71	0.71	0.73	0.69
13	0.68	0.76	0.77	0.69	0.55	0.50	0.60	0.67	0.68	0.67	0.68	0.68
14	0.68	0.74	0.80	0.69	0.58	0.50	0.53	0.59	0.64	0.63	0.73	0.65
15	0.66	0.67	0.80	0.60	0.53	0.63	0.55	0.57	0.63	0.63	0.64	0.64
16	0.66	0.67	0.67	0.62	0.60	0.52	0.53	0.53	0.57	0.65	0.65	0.61
17	0.64	0.69	0.67	0.74	0.63	0.37	0.39	0.54	0.59	0.65	0.66	0.70
18	0.68	0.66	0.66	0.66	0.56	0.40	0.35	0.41	0.56	0.61	0.58	0.70
19	0.65	0.67	0.63	0.61	0.53	0.52	0.39	0.36	0.42	0.60	0.58	0.61
20	0.73	0.68	0.67	0.60	0.61	0.40	0.34	0.30	0.34	0.36	0.58	0.61
21	0.64	0.57	0.56	0.56	0.46	0.49	0.33	0.34	0.34	0.33	0.35	0.57
22	0.53	0.58	0.54	0.51	0.46	0.40	0.36	0.30	0.32	0.30	0.34	0.36
23	0.43	0.53	0.58	0.48	0.38	0.38	0.39	0.32	0.37	0.36	0.35	0.29

Table S2. Correlation skill - CNN

	Target season											
Forecast lead	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF
1	0.96	0.95	0.91	0.89	0.92	0.88	0.93	0.94	0.97	0.96	0.96	0.96
2	0.94	0.94	0.90	0.85	0.84	0.84	0.91	0.93	0.97	0.96	0.95	0.95
3	0.92	0.91	0.89	0.79	0.75	0.77	0.83	0.90	0.93	0.95	0.95	0.94
4	0.90	0.89	0.83	0.75	0.68	0.68	0.75	0.81	0.90	0.92	0.95	0.93
5	0.91	0.86	0.82	0.71	0.63	0.66	0.68	0.74	0.81	0.91	0.92	0.92
6	0.90	0.88	0.83	0.69	0.55	0.60	0.68	0.65	0.74	0.83	0.90	0.90
7	0.93	0.88	0.84	0.73	0.56	0.58	0.59	0.65	0.65	0.76	0.82	0.87
8	0.88	0.90	0.86	0.72	0.60	0.61	0.65	0.66	0.69	0.66	0.73	0.80
9	0.82	0.85	0.84	0.73	0.60	0.66	0.63	0.71	0.65	0.68	0.66	0.74
10	0.80	0.80	0.80	0.65	0.60	0.60	0.66	0.68	0.72	0.67	0.67	0.68
11	0.73	0.76	0.79	0.62	0.50	0.59	0.59	0.64	0.70	0.69	0.68	0.70
12	0.77	0.79	0.79	0.61	0.40	0.56	0.58	0.62	0.66	0.68	0.70	0.65
13	0.73	0.75	0.76	0.68	0.41	0.40	0.56	0.62	0.69	0.71	0.64	0.65
14	0.66	0.65	0.73	0.65	0.44	0.43	0.48	0.55	0.64	0.67	0.66	0.61
15	0.63	0.59	0.70	0.56	0.44	0.40	0.46	0.47	0.58	0.66	0.67	0.66
16	0.65	0.60	0.63	0.53	0.33	0.30	0.40	0.48	0.51	0.62	0.68	0.63
17	0.66	0.64	0.55	0.55	0.41	0.29	0.27	0.41	0.48	0.50	0.65	0.67
18	0.64	0.69	0.56	0.52	0.37	0.26	0.22	0.24	0.37	0.48	0.54	0.63
19	0.60	0.66	0.63	0.52	0.27	0.29	0.22	0.20	0.27	0.41	0.45	0.57
20	0.49	0.60	0.63	0.47	0.37	0.26	0.22	0.21	0.22	0.23	0.42	0.45
21	0.39	0.48	0.55	0.50	0.31	0.24	0.23	0.24	0.25	0.23	0.27	0.43
22	0.38	0.38	0.50	0.45	0.29	0.28	0.24	0.17	0.26	0.26	0.23	0.33
23	0.28	0.38	0.40	0.44	0.27	0.17	0.25	0.27	0.28	0.22	0.25	0.24

Table S3. Correlation skill - SINTEX-F

	Target season											
Forecast lead	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF
1	0.95	0.95	0.93	0.80	0.71	0.73	0.87	0.91	0.93	0.96	0.96	0.97
2	0.94	0.93	0.88	0.71	0.71	0.78	0.87	0.92	0.93	0.95	0.96	0.96
3	0.93	0.90	0.82	0.70	0.61	0.73	0.79	0.87	0.91	0.92	0.94	0.93
4	0.91	0.88	0.80	0.59	0.56	0.65	0.75	0.80	0.87	0.92	0.92	0.93
5	0.92	0.87	0.81	0.58	0.40	0.62	0.69	0.74	0.78	0.88	0.93	0.91
6	0.91	0.89	0.80	0.60	0.36	0.42	0.71	0.66	0.73	0.78	0.89	0.92
7	0.92	0.88	0.81	0.61	0.40	0.34	0.54	0.71	0.61	0.72	0.79	0.89
8	0.90	0.88	0.79	0.57	0.41	0.38	0.44	0.60	0.69	0.59	0.72	0.79
9	0.83	0.86	0.78	0.54	0.34	0.39	0.47	0.50	0.61	0.67	0.60	0.71
10	0.81	0.80	0.76	0.54	0.30	0.29	0.50	0.53	0.51	0.59	0.67	0.61
11	0.68	0.78	0.71	0.53	0.30	0.25	0.36	0.58	0.54	0.50	0.58	0.67
12	0.72	0.63	0.72	0.49	0.30	0.24	0.35	0.43	0.61	0.53	0.52	0.55
13	0.55	0.67	0.55	0.55	0.27	0.23	0.31	0.43	0.45	0.62	0.54	0.55
14	0.63	0.52	0.58	0.44	0.40	0.19	0.32	0.40	0.47	0.44	0.64	0.55
15	0.54	0.62	0.45	0.42	0.30	0.31	0.17	0.42	0.44	0.46	0.47	0.65
16	0.65	0.49	0.57	0.34	0.24	0.24	0.31	0.18	0.47	0.45	0.49	0.51
17	0.54	0.58	0.41	0.40	0.24	0.18	0.23	0.35	0.16	0.47	0.47	0.52
18	0.54	0.55	0.46	0.29	0.20	0.23	0.18	0.25	0.37	0.15	0.48	0.49
19	0.54	0.52	0.51	0.30	0.20	0.11	0.26	0.21	0.24	0.38	0.17	0.48
20	0.47	0.52	0.42	0.43	0.19	0.24	0.11	0.28	0.23	0.22	0.40	0.21
21	0.25	0.43	0.46	0.25	0.33	0.21	0.35	0.14	0.30	0.26	0.24	0.44
22	0.50	0.25	0.38	0.36	0.13	0.30	0.28	0.42	0.16	0.29	0.31	0.26
23	0.24	0.50	0.26	0.34	0.25	0.15	0.32	0.34	0.44	0.18	0.30	0.35