



## OPEN Taking advantage of quasi-periodic signals for S2S operational forecast from a perspective of deep learning

Yang Zhou<sup>1,2✉</sup> & Qifan Zhao<sup>1</sup>

The quasi-periodic signals in the earth system could be the predictability source for sub-seasonal to seasonal (S2S) climate prediction because of the connections among the lead-lag time of those signals. The Madden–Julian Oscillation (MJO) is a typical quasi-periodic signal, which is the dominant S2S variability in the tropics. Besides, significantly periodic features in terms of both intensity and location are identified in 10–40 days for the concurrent variation of the subtropical and polar jet streams over Asia in this study. So far, those signals contribute less and are not fully applied to the S2S prediction. The deep learning (DL) approach, especially the long-short term memory (LSTM) networks, has the ability to take advantage of the information at the previous time to improve the prediction after then. This study presents the application of the DL in the postprocessing of S2S prediction using quasi-periodic signals predicted by the operational model to improve the prediction of minimum 2-m air temperature over Asia. With the help of deep learning, it finds the best weights for the ensemble predictions, and the quasi-periodic signals in the atmosphere can further benefit the S2S operational prediction.

The numerical climate prediction on sub-seasonal to seasonal (S2S) time scales has been operational in weather/climate forecast centers worldwide for years<sup>1</sup>. Compared to the short-range weather forecast, its extended range of lead time can provide valuable hours for thoughtful decisions and adequate preparations, but skillful S2S prediction with a long lead time is a remaining challenge for meteorologists<sup>2,3</sup>. As an important criterion to assess S2S operational models, a lot of effort has been invested in the improvement of predicting quasi-periodic signals such as the Madden–Julian Oscillation (MJO), which is the dominant S2S signal in the tropical atmosphere<sup>4–6</sup>. In the operational models, the lead time of skillful MJO prediction can be over 30 days<sup>7,8</sup>, which is much longer than those for the other atmospheric variables in extratropical areas such as 2-m air temperature<sup>9</sup>. Though MJO significantly influences the S2S prediction, the improvement of MJO prediction brings limited benefit for S2S prediction about the other atmospheric variables<sup>9–11</sup>. This is partly because the numerical model may not well express the complex interactions between MJO and the atmosphere. Moreover, the effects of MJO on S2S prediction are not extensively revealed and mainly derived from traditional statistical methods, which are generally restricted to linear approaches. While numerical models are developed for improving S2S prediction, how can we seek a way to take advantage of the quasi-periodic signals (e.g., MJO) in S2S prediction?

Nowadays the major tool for S2S prediction (between two weeks and a season) is the dynamic model, which is driven by both initial and boundary conditions. Forecasts of less than two weeks depend on initial conditions<sup>12</sup>, and the slow-evolving aspects incorporated in the boundary conditions are dominant factors for climate predictions longer than a season<sup>13</sup>. For the S2S prediction, the model loses most of the memory of its initial state, and the time period is too short for the model atmosphere to acquire notable responses to the slow-evolving parts of the boundary conditions. Thus, the variabilities in the atmosphere on S2S time scales become an important source of predictability<sup>14–16</sup>. For example, MJO impacts the evolution of midlatitude circulation several weeks later<sup>17–19</sup>, and significant improvement in MJO prediction is achieved in the S2S operational models<sup>7,8</sup>. The Beijing Climate Center (BCC) model, which is the S2S operational model of the Chinese Meteorology Administration (CMA) joining the S2S project (<http://www.s2sprediction.net>), can provide skillful MJO prediction about 20 days ahead<sup>20,21</sup>, but the other variables such as 2-m air temperature in the model still have the short lead time. Besides prediction itself, S2S climate prediction is an issue with the postprocessing of massive model outputs<sup>1</sup>.

<sup>1</sup>Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China. <sup>2</sup>School of Atmospheric Sciences, Nanjing University of Information Science and Technology, No. 219 Ningliu Road, Pukou District, Nanjing 210044, Jiangsu, China. ✉email: yangzhou@nuist.edu.cn

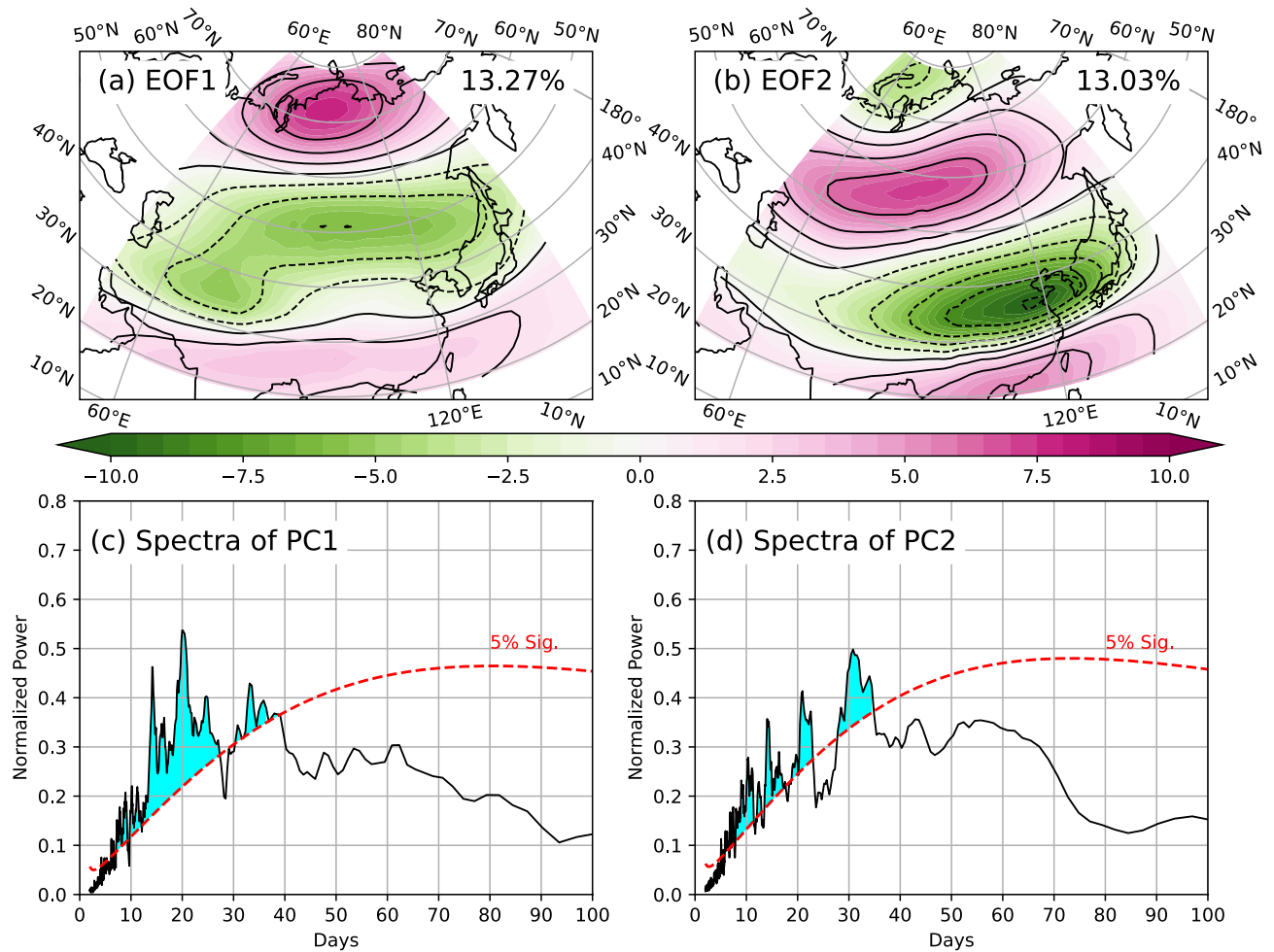
The major task of postprocessing is to achieve skillful S2S prediction, which can adopt deep learning (DL)<sup>22</sup>. DL is used to postprocess outputs from ensemble members of S2S forecasts to obtain better predictions of MJO and atmospheric variables than those of the direct ensemble mean<sup>23,24</sup>.

Confronting massive atmospheric datasets and complex nonlinear problems, efficient statistical tools are required beyond linear approaches to express the nonlinearity and complement current understanding. Nowadays state-of-the-art numerical models help understand and predict the climate system much better than before, but complex phenomena and useful information still need to be dug in numerous datasets yielded year on year<sup>25</sup>. Although the models are based on nonlinear dynamics, analysis methods for the model outputs are mostly linear statistics<sup>26,27</sup>, which to a certain extent preclude most of the nonlinearity. According to the outstanding performance in image processing and language recognition, DL becomes a powerful engine for massive data analysis with nonlinearity being considered<sup>28,29</sup>. Machine learning (ML) and DL have become modern tools in earth science with large amounts of observation and model data<sup>30–32</sup>. Due to the success in dealing with images, DL is applied in the identification of weather patterns<sup>33,34</sup>, the category of precipitation types<sup>35</sup>, and the classification of the wave breaking in the ocean<sup>36</sup>. Based on satellite and radar images, DL is further used in estimations of rainfall and tropical cyclone intensity<sup>37–39</sup>. DL is also employed as a useful tool for forecasting air temperature, rainfall, and air quality<sup>40–43</sup>. Though based on a purely data-driven approach, DL can even simulate physical laws. Some studies tried to substitute DL for the physical models in weather forecasts, and DL performed much better than some simple physical models, but it has not beaten the operational, dynamical models yet<sup>44–46</sup>. Furthermore, an approach of physics-informed ML suggests guiding ML with known physical laws to improve the forecasts<sup>47</sup>. Meanwhile, the hybrid approach combining ML with physical modeling is the most applied one<sup>25</sup>. For example, DL is used for data assimilation and moist convection parameterization in climate models<sup>48–50</sup>. DL is applied for postprocessing of model outputs to remove systematic errors and allow the nonlinearity among ensemble members of the models<sup>24,51,52</sup>. Driven by the output data of the dynamical models, DL has yielded promising climate predictions for the East Asian Summer monsoon and El Niño Southern Oscillation<sup>53–55</sup>.

Minimum 2-m air temperature (T2Min) is one of the important variables for identifying extreme weathers in winter<sup>56,57</sup>. The cold events are usually defined based on the T2Min values<sup>58</sup>. Thus, the skillful prediction of T2Min is substantial for the prediction of extreme cold weathers. Changes in the position and strength of jet streams can significantly affect the atmospheric circulation and storm tracks over the mid-latitude Northern Hemisphere in winter, which can alter the weather patterns<sup>59–63</sup>. Therefore, the variations of the jet streams have significant influences on the winter air temperature over Asia<sup>64–66</sup>. Furthermore, jet streams are one of the major linkages between severe winters over the Northern Hemisphere and other climatic factors<sup>67,68</sup>. Recently, a study has also pointed out that jet stream and MJO have the joint influence on the S2S winter patterns over the Northern Hemisphere<sup>69</sup>. Generally, MJO is a useful signal with a long lead time of prediction in the model. Due to the quasi-periodic nature of MJO, the atmosphere can be influenced by MJO several weeks later, which indicates these kinds of signals can be used for prediction. The jet streams are important factors affecting winter temperature over Asia, and the present study is going to reveal that the concurrent variation of the subtropical and polar jet streams is another quasi-periodic signal over the midlatitude of Asia. On the other hand, DL is a useful tool for the postprocessing of model outputs, which can incorporate nonlinearity and even physical laws, though it is a black box that is difficult to track any physical process. Since MJO is well predicted by the model and MJO and jet streams have quasi-periodic nature, DL can be used to combine these kinds of signals and the atmospheric variables in postprocessing model outputs to improve the S2S prediction. In other words, we can use DL to transfer the benefit from the prediction of quasi-periodic signals to the prediction of the variables cared by the public. Based on this idea, a postprocessing approach of DL is proposed for combining the quasi-periodic signals and T2Min predicted by the operational S2S model of CMA (BCC model) to improve the S2S prediction of T2Min over Asia in boreal winter. This can potentially contribute to the S2S prediction of severe cold weathers over Asia.

## Results

**Quasi-periodic features of jet and MJO.** In order to explore the quasi-periodic features of jet streams, EOF analysis as described in Methods is conducted on daily  $U_{\text{ERA}}$  at 200 hPa over Asia during winter (Fig. 1). The first EOF (EOF1) explains 13.27% of the total variances and shows a tripole pattern (Fig. 1a). A positively anomalous center of the  $U_{\text{ERA}}$  is over the polar region around 70°N. A negatively anomalous band is around 50°N (near the polar jet stream), and a weak positively anomalous band is around 25°N (around the subtropical jet stream). EOF1 generally reflects that when the zonal wind anomalies are strong and westerly wind increases over the polar region, the polar jet stream weakens and the subtropical jet stream enhances, and vice versa. Moreover, intensities of the polar and subtropical jet streams present opposite or seesaw variations. The second EOF (EOF2) explaining 13.03% of the total variances mainly presents a seesaw pattern over Asia (Fig. 1b), and two strong anomalous bands are around 55°N and 35°N with opposite signs. This pattern also presents the opposite variations of intensities between the polar and subtropical jet streams. Compared with EOF1 (Fig. 1a), the intensity of the zonal wind anomalies of EOF2 (Fig. 1b) is strong, and the locations of the seesaw anomalous bands are further north. Besides, the intensity of zonal wind over the polar region is weak, and there is a positively anomalous band around 20°N (Fig. 1b). The first two EOFs indicate a concurrent variation of both intensity and location of the two jet streams on S2S time scales over Asia during winter. To further explore features of this variation, spectral analysis is conducted on the time series of the two PCs (JET1&2) and shown in Figs. 1c and d. Significant spectra are found during 10–40 days for both JET1 and JET2. The significant spectral peaks are on 10, 15, 20, and 35 days for JET1 (Fig. 1c) and on 10, 15, 20, and 30 days for JET2 (Fig. 1b). Therefore, the concurrent variation of the two jet streams has significant oscillation on S2S time scales, and the bivariate indices

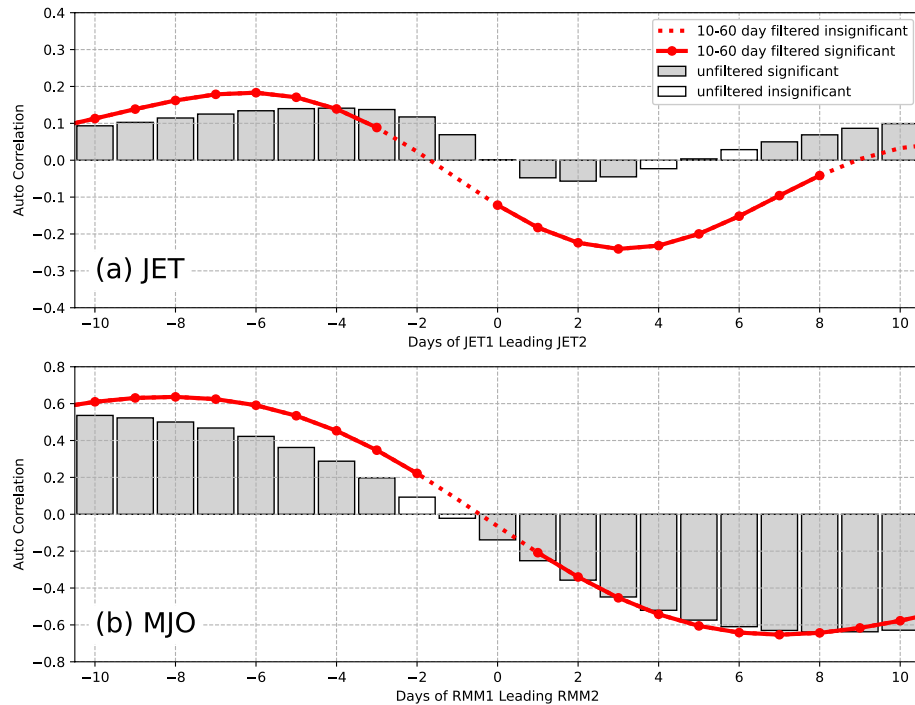


**Figure 1.** First two EOFs (panels **a** and **b**) of the daily  $1.5^\circ \times 1.5^\circ U_{\text{ERA}}$  at 200 hPa during boreal winter. Before the EOF, the seasonal cycle has been removed. The first EOF (EOF1) explains 13.27% of the total variance, and the second (EOF2) explains 13.03%. EOF is conducted over the region of  $49.5\text{--}148.5^\circ\text{E}$  and  $19.5\text{--}79.5^\circ\text{N}$  (shaded area). The two panels share the same color bar. Power spectra (black curve) for the two PCs of the first two EOFs (JET1&2) are shown in panels **c** and **d**, respectively. The red-dash line shows the 5% significant level based on the Markov red noise, and the significant power spectra are shaded in blue. The x-axis is the period, and the y-axis is the normalized power spectral.

of jet streams also present the propagation of those S2S signals. For the MJO, its convection mainly propagates eastwardly in the tropics, which was explored by many previous studies<sup>4–6,70</sup>.

To further reveal the quasi-periodic nature of jet and MJO, Fig. 2 shows the lead-lag correlation coefficients between JET1 and JET2 during the winters of 1979–2019, as well as those for the MJO indices of RMM1 and RMM2. Due to those oscillations being on S2S time scales, band-pass filtering with the cutoff of 10–60 days using the Butterworth filter is also conducted on the time series of those indices. The correlation coefficients are generally weak and insignificant around day zero because of the orthogonal feature of the PCs of EOF analysis. For the jet stream (Fig. 2a), when JET1 leads JET2, the largest correlation coefficients are positive on days 4 and 6 for unfiltered and filtered data, respectively. The positive correlation indicates EOF1 leads EOF2, and vice versa. When JET2 leads JET1, the largest correlation coefficients are negative on days 2 and 3 for unfiltered and filtered data, respectively. For MJO (Fig. 2b), when RMM1 leads RMM2, the largest correlation coefficients are on days 10 and 8 for unfiltered and filtered data, respectively. When RMM2 leads RMM1, the largest correlation coefficients are on days 8 and 7 for unfiltered and filtered data, respectively. Except around day zero, both the unfiltered and filtered data generally exhibit significant lead-lag correlations between the two components of the bivariate indices of either MJO or jet. This indicates that the first two EOF patterns of either jet and MJO substitute each other on S2S time scales as time varying. Thus, each component of either indices of those two quasi-periodic signals incorporates future information of the other component for about 4–10 days. Overall, the two indices of MJO and jet have quasi-periodic features both spatially and temporally and their signals are time lagged, which can be applied for the usage of prediction.

The linear correlations between RMMs and JETs are insignificant, and there are significant but weak linear correlations between RMMs/JETs and T2Min, except JET2 (figures not shown), which is mainly because the



**Figure 2.** (a) Lead-lag correlation coefficients between JET1 and JET2 during the winters of 1979–2019, as well as those between RMM1 and RMM2 (b). The correlation coefficients for the unfiltered data are shown in bars, and those for 10–60-day filtered data are shown in red lines. The significant correlation coefficients at the 5% level are presented by filled bars or solid lines.

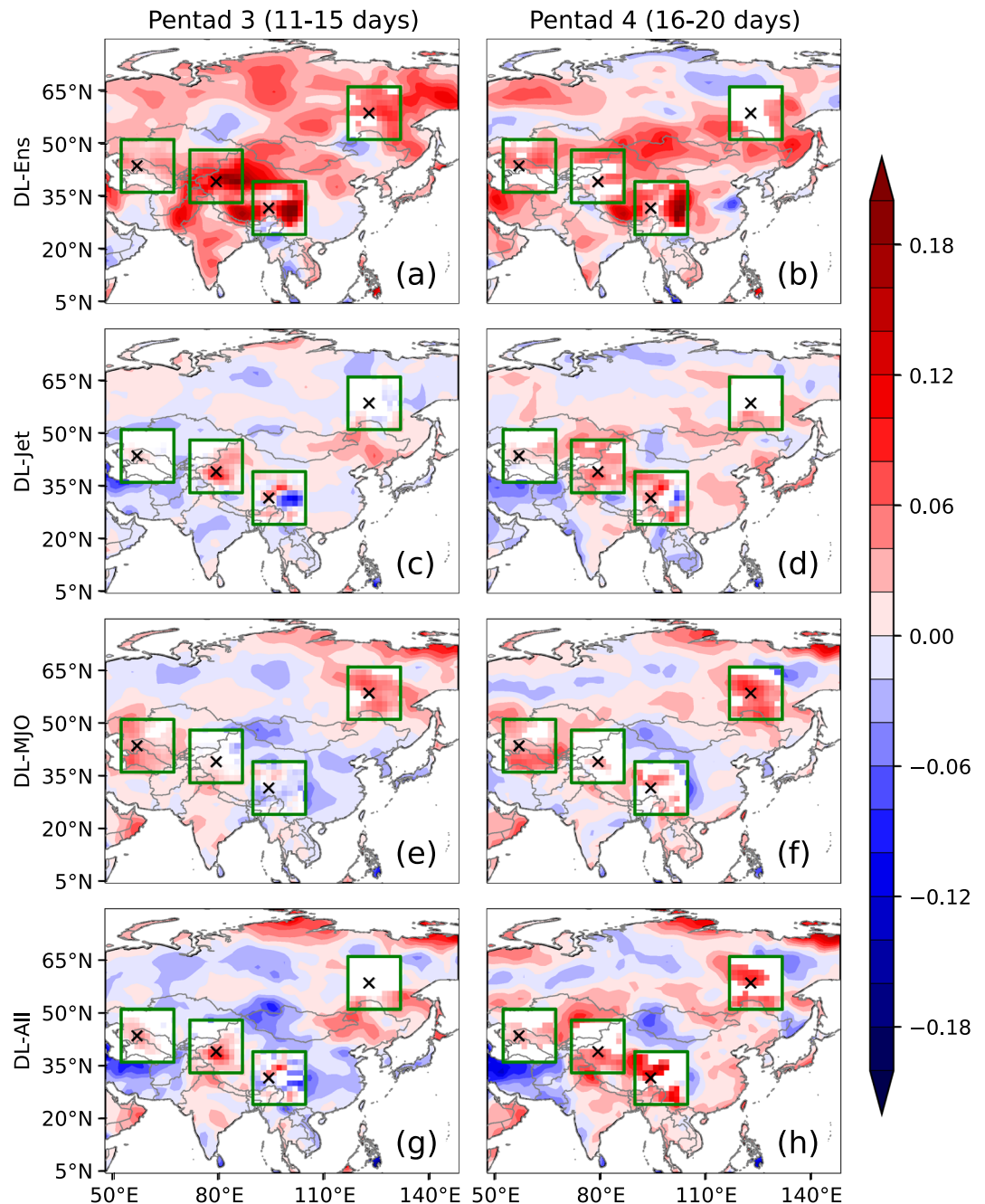
significant influences of MJO and jet streams on air temperature over Asia<sup>18,64–66</sup>. In the following, RMMs and JETs are used as input features for DL models.

**Postprocessing by the DL models over Asia.** Using the T2Min, JET1, JET2, RMM1, and RMM2 from the S2S products of the BCC model, four DL models (Table 1 in “Methods” section) are trained against the T2Min of ERA5 during winters of 2004–2014 (10 winters). The four models have input features from 4 to 20 with a sequence length of 30 and batch size of 26 ( $30 \times 26 \times N$ ,  $N$  indicates the size of input features). The output of a model is a matrix of  $30 \times 26$  of T2Min, which are the sequence length and batch size, respectively. The sequence length is the length of forecast days, and batch size is the number of forecasts in a winter. Predictions of the DL models are validated during the winters of 2015–2019 (5 winters). The BCC model generally has skillful prediction during the first 10 days for T2Min, and thus the improvement of correlation skill (R) and RMSE for the four DL models over Asia at pentad 3 (11–15 forecast days) and 4 (16–20 forecast days) is shown in Figs. 3 and 4. The prediction of T2Min is averaged at pentad 3 and 4, and then R and RMSE are calculated against T2Min of ERA5.

Figures 3a and b show the differences in the skill between DL-Ens and the direct mean of ensemble members ( $\Delta R_{DL-Ens}$ ). Compared with the direct mean of the ensemble members, DL-Ens improves the prediction skill of T2Min over most parts of Asia. At pentad 3 (Fig. 3a), the improvement is found over most parts of Russia, southern Kazakhstan, Mongolia, western China, Pakistan, and India. At pentad 4 (Fig. 3b), areas with the improvement are similar to those at pentad 3, except northern Russia and part of central Asia. At both pentad 3 and 4, the largest improvement is over western China, and the prediction skill is generally worsened over northern Kazakhstan and southeastern China. In order to check whether jet and MJO can improve the prediction skills, the differences in the prediction skills between DL-jet and DL-Ens are presented in Figs. 3c and d, as well as those between DL-MJO/DL-All and DL-Ens (Figs. 3e–h). When the jet signal predicted by the BCC model is

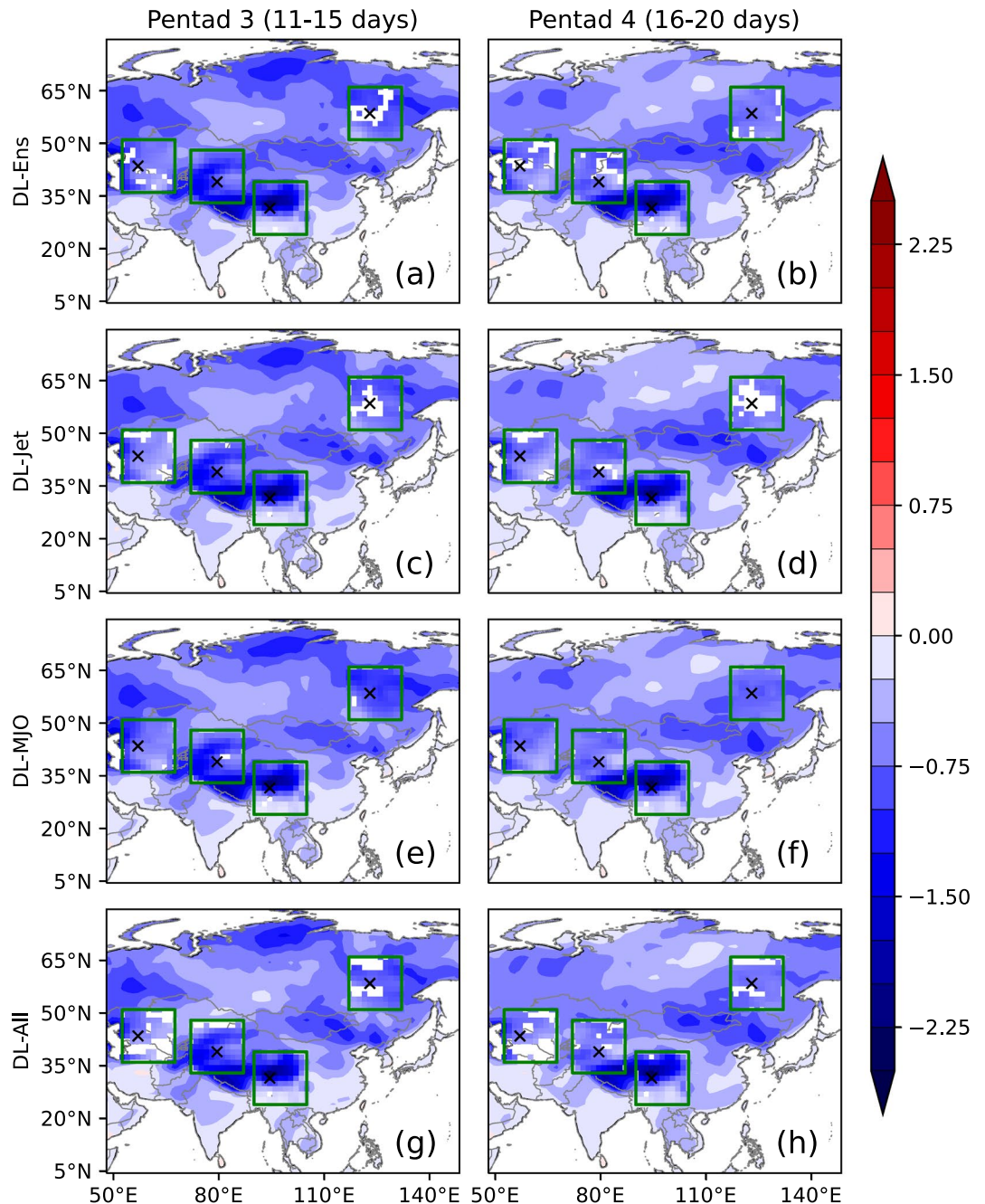
| DL Model Name | Number of Input Features               | Features                  |
|---------------|--|---------------------------|
| DL-Ens        | $N = 4$                                | T2Min                     |
| DL-Jet        | $N = 4 + 2 \times 4 = 12$              | T2Min and JET1&2          |
| DL-MJO        | $N = 4 + 2 \times 4 = 12$              | T2Min and RMM1&2          |
| DL-All        | $N = 4 + 2 \times 4 + 2 \times 4 = 20$ | T2Min, JET1&2, and RMM1&2 |

**Table 1.** Four DL models with different numbers of features (N) for the input layer. For the BCC model, each of T2Min, RMM1, RMM2, JET1, and JET2 at a grid cell has the predictions of four ensemble members.



**Figure 3.** Differences of the correlation skill ( $R$ ) among the direct mean of ensemble members and the four DL models at pentad 3 (left panels) and 4 (right panels). (a and b) Differences between  $R_{DL-Ens}$  and the direct mean of ensemble members. (c and d) Differences between  $R_{DL-Jet}$  and  $R_{DL-Ens}$ . (e and f) Differences between  $R_{DL-MJO}$  and  $R_{DL-Ens}$ . (g and h) Differences between  $R_{DL-All}$  and  $R_{DL-Ens}$ . Due to the large consumption of the computation resources for the significant test, four regions surrounded by green rectangles are chosen for the significant test, and the grid cells filled with color are significant at the 5% level. Those regions are denoted as REG1, 2, 3, and 4 from left to right. The cross markers are used for Figs. 5 and 6. The DL models are built based on the pytorch module (version 1.10.0, <https://pytorch.org/>) of python. The figure is drawn using the modules of matplotlib (version 3.4.1, <https://matplotlib.org/>) and cartopy (version 0.18.0, <https://scitools.org.uk/cartopy/>) of python.

considered in the DL-Jet, compared to DL-Ens, the prediction is improved over western and northeastern China at pentad 3 (Fig. 3c) and eastern Kazakhstan, western and eastern China, and southeastern Russia at pentad 4 (Fig. 3d). When the MJO predicted by the BCC model is considered in the DL-MJO, compared to DL-Ens, the prediction skill is increased over central Asia, part of western China, and eastern Russia at both pentad 3 and 4 (Figs. 3e and f). When both the jet and MJO predicted by the BCC model are used in DL-All, prediction is

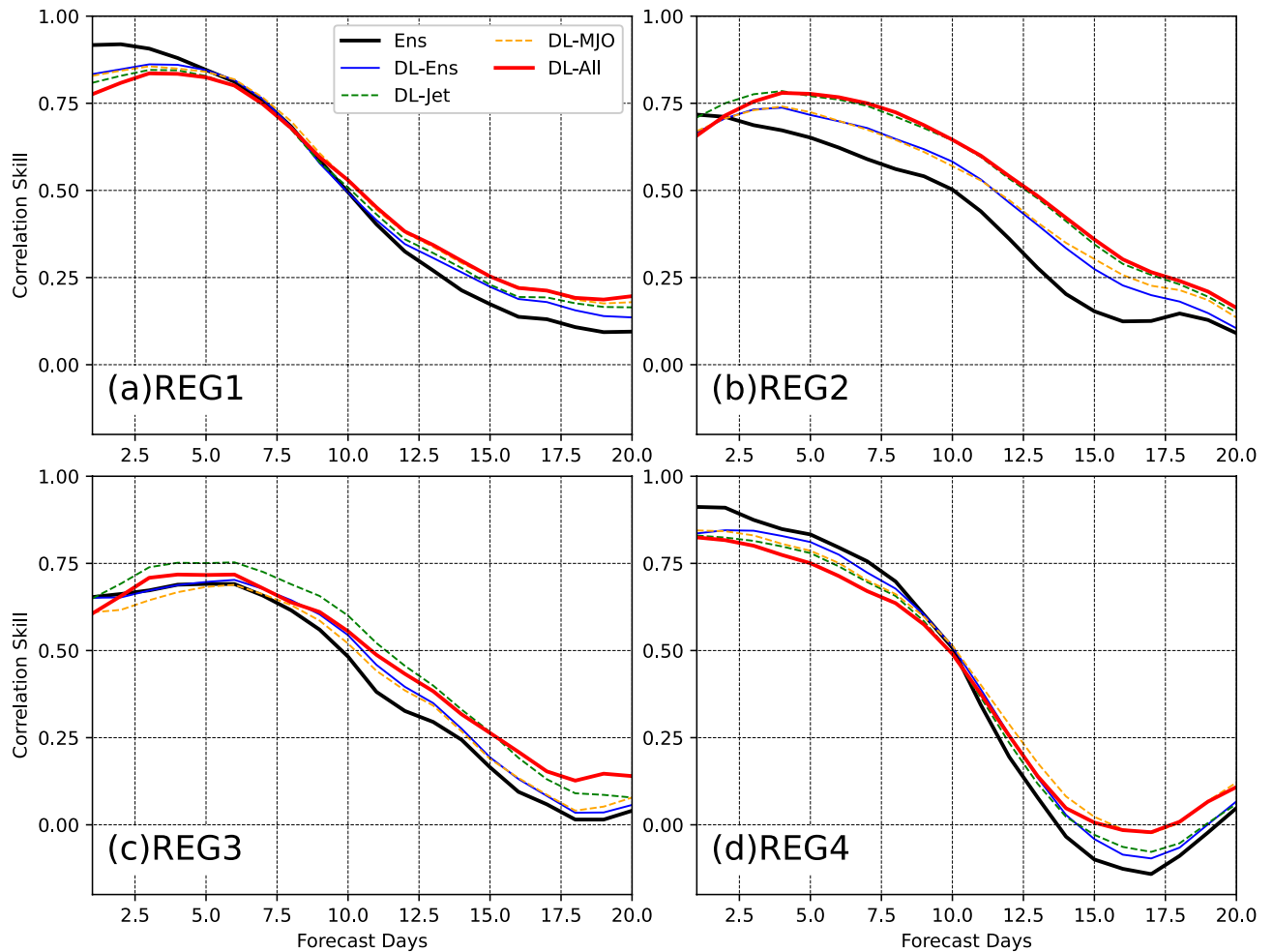


**Figure 4.** Same as Fig. 3, but for the differences of RMSE.

improved over western and northeastern China at pentad 3 (Fig. 3g) and over Kazakhstan, western China, and eastern Russia at pentad 4 (Fig. 3h).

In the regions surrounded by green rectangles (Fig. 3), the significant test is conducted for the improvement of the prediction skill at each grid cell, and significant improvement is filled with color. For the color-filled grid cells over the four regions, the prediction skill is generally significantly improved (red) at pentad 3 and 4, except for some grid cells over central China, where the prediction skill is significantly decreased (blue). Contrary to those areas with improved prediction skills, some regions are with a decrease of the prediction skill postprocessed by the DL models. Thus, the applicability of DL method exhibits some regional dependency.

The same as Fig. 3, the RMSE at pentad 3 and 4 is shown in Fig. 4. It is noted that the seasonal mean and mean of both model results and observations are removed before the calculation of RMSE, and thus RMSE is unbiased. The RMSE is decreased in DL-Ens compared with the direct mean of ensemble members. When the jet and MJO predicted by the BCC model are considered, the RMSE at pentad 3 and 4 in DL-Jet, DL-MJO, and DL-All is also decreased compared with DL-Ens. In Fig. 4, RMSE is either decreased or unchanged but not increased, and the largest decrease of RMSE is over the Tibetan Plateau. In the four regions surrounded by the green lines, most



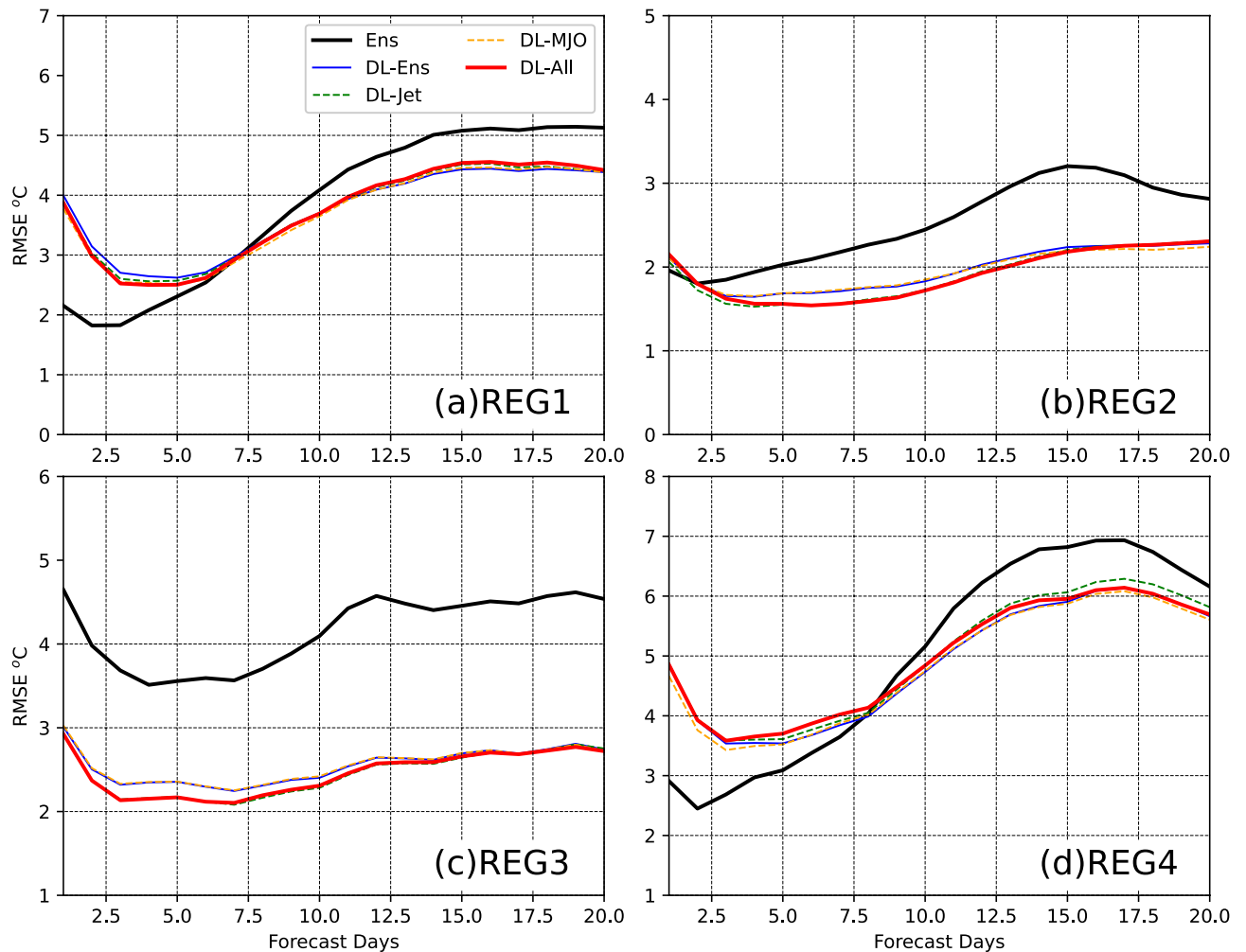
**Figure 5.** Time variations of R at four grid cells in the four regions marked in Figs. 3 and 4. The R values are presented during the first 20 forecast days. The solid black curve is the R for the direct mean of ensemble members. The solid blue, dash green, dash orange, and solid red lines are for DL-Ens, DL-Jet, DL-MJO, and DL-ALL, respectively. The y-axis and x-axis are for R and forecast days, respectively.

parts are with a significant decrease in RMSE. Therefore, the DL method can significantly decrease the error of the BCC model, though it cannot increase the prediction skill over entire Asia.

**Time variations of prediction skill of the DL models.** To further compare the DL models and present time variations of the prediction skill, the time variations of R and RMSE during the first 20 forecast days are presented in Figs. 5 and 6. The R and RMSE are calculated at the four grid cells in the four regions shown by the cross markers in Figs. 3 and 4.

For the BCC model, the direct mean of the ensemble members has the prediction skill greater than 0.5 during the first 10 forecast days at the four points (Fig. 5).  $R \geq 0.5$  indicates the skillful prediction of T2Min. During the first 10 forecast days, the prediction skill in REG2 and 3, where are western China and the eastern Tibetan Plateau, are smaller than that in REG1 and 4, where are central Asia and eastern Russia. After 10 days, the prediction skill is not skillful at all four points, and thus the BCC model generally has the ability to the prediction T2Min with a lead time of 10 days. After postprocessing by the DL model, the prediction skill has been improved after 10 forecast days. During the first 10 forecast days, the skill is decreased at the points in REG1 and 4 but increased in REG2 and 3 in DL-models.

Because the prediction during the first 10 forecast days is skillful in the BCC model, we mainly focus on the improvement of the prediction after 10 forecast days. In REG1 (Fig. 5a), the prediction skill is improved by all the DL models after 10 forecast days, and the improvement of DL-All is the largest among them with the improvement of lead days of about 0.5. DL-Jet and DL-Ens almost have the same improvement, and DL-MJO and DL-All almost overlap with each other. This indicates that MJO contributes more to the improvement by the postprocessing at the point in REG1. In REG2 (Fig. 5b), DL-All also has the largest improvement among the forecasts with the improvement of lead days up to 2.5 days. DL-MJO and DL-Ens have the similar improvement, and DL-All and DL-Jet almost overlap with each other. Thus, jet has larger contribution to the postprocessing than MJO at the point in REG2. In REG3 (Fig. 5c), the prediction of DL-Jet has the largest skill during 10–15 days among the forecasts, but that of DL-All has the largest during 15–20 days. The jet contributes more to the prediction skill



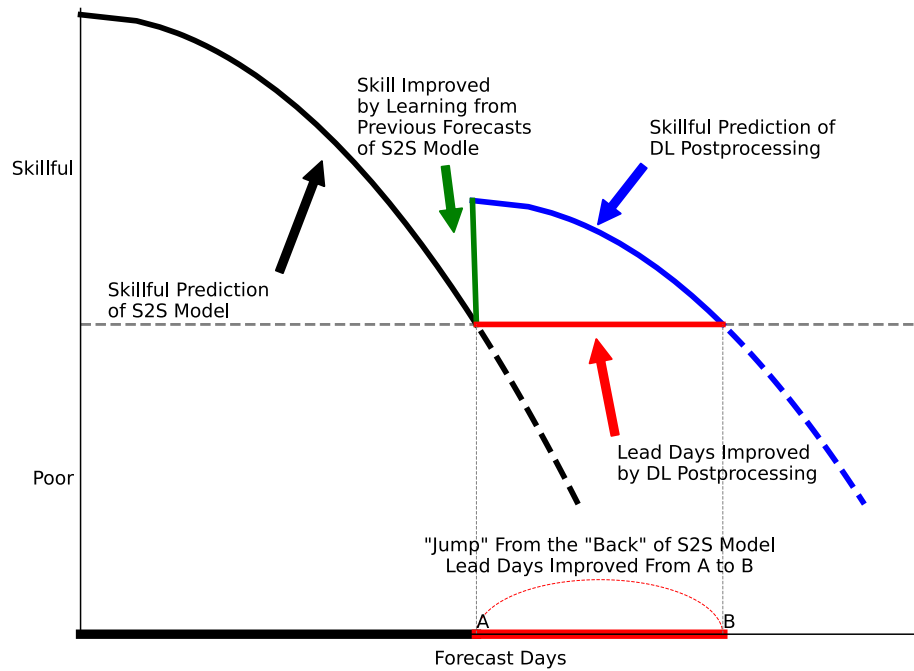
**Figure 6.** Same as Fig. 5, but for RMSE.

than MJO at the point in REG3. In REG4 (Fig. 5d), DL-Ens, DL-Jet, and DL-All almost have the same improvement of the skill during 10–15 days. During 15–20 days, DL-All and DL-MJO have the largest improvement. At this point in REG4, MJO contributes more to the prediction than the jet. Similar results can be found at most of the points in the four regions, which can be also derived from Fig. 3. When considering MJO predicted by the BCC model in the DL postprocessing, the prediction skill is improved over central Asia and eastern Russia. When considering jet predicted by the BCC model in the DL postprocessing the skill is improved over western China and the eastern Tibetan Plateau. This further proved that the applicability of DL method exhibits some regional dependency. For the RMSE (Fig. 6), the four DL models present negligible differences and similar variations during the 20 forecast days. At the points in REG1 and 3, the RMSE is increased by the DL models before 7.5 forecast days but decreased after then. At the points in REG2 and 3, the RMSE is generally decreased during 20 forecast days.

Overall, the DL postprocessing can generally increase the prediction skill and decrease the RMSE compared to the direct mean of ensemble members over Asia. Quasi-periodic signals (e.g., jet and MJO) cannot improve the prediction everywhere over Asia but can improve the skill over some specific regions.

**Discussions.** The T2Min in the BCC model generally has skillful prediction during the first 10 forecast days, after then the skill drops. In the present study, DL can find the best weights for the ensemble members to significantly improve the prediction during postprocessing. Like many DL postprocessing, this approach only uses the predictions of T2Min itself, because the nonlinearity and sequence dependence are considered by the LSTM model. Besides the prediction of T2Min itself, there are many other factors predicted by the model, which have better prediction skills or quasi-periodic nature. Moreover, those factors can influence the T2Min physically in the observation, and many previous studies have illustrated that MJO and jet stream have significant effects on the midlatitude climate<sup>4–6,9,17,18,71,72</sup>. MJO and jet have quasi-periodic features, and the model can predict them well in the first 10 forecast days. Thus, using the DL approach, the information of MJO and jet that are well predicted in the previous time can be used to correct the prediction after then according to the quasi-periodic nature of the signals (Fig. 7). Moreover, the prediction skill of MJO is higher than that of T2Min in the BCC model, and those skills can be also used by the DL approach to improve the prediction of T2Min, though it is





**Figure 7.** Schematic diagram for the DL approach improving the prediction skill of S2S operational models.

difficult to separate this effect from other effects in the DL models. The DL approach in this study suggests using the information that is well predicted in the first two weeks of the forecast to improve the forecast after then, based on the quasi-periodic feature of that information (Fig. 7). It is like an approach to “jump” from the “back” of S2S model to improve the lead days of the prediction of T2Min. Thus, with the improvement of the numerical model, the DL postprocessing could always have a better prediction than the numerical model itself.

It is interesting to find that the prediction skill at pentad 4 is improved more than that at pentad 3 in the DL models. We think this probably because the useful information can be provided by the LSTM model is already provided by the numerical model at pentad 3. Thus, less useful signal is provided by the LSTM model for improvement. At pentad 4, the numerical model losses more information from the initial condition than at pentad 3. When the prediction skill in the numerical model drops faster than the memory of the LSTM model, the LSTM model can provide more useful information than that at pentad 3. Therefore, the prediction skill is improved more at pentad 3 than at pentad 4.

It is also noted that the LSTM network does not characterize the spatial motion of the atmosphere. LSTM is chosen because we mainly want to get the prediction at each grid (or location) that is only the result of the signals in jet/MJO but not the atmospheric variables on other grids. Although the spatial motion of the atmosphere over the grid is the reason for the effects of jet/MJO on prediction, it is not considered in the present study. On the other hand, the dynamic model already includes the results of the spatial motion of the atmosphere. This part maybe not well captured by the dynamic model, and a DL model has the potential to improve it. However, it is another interesting issue that can be further explored.

The DL-Ens is an approach to find the best weights for the ensemble members for the S2S products, and prediction skill is significantly improved over most parts of Asia. This indicates that the direct mean of the ensemble members (equal weights) can be improved, though its prediction skill is better than any single member. DL-Jet, DL-MJO, and DL-All consider the signals of jet and MJO during the postprocessing, and the prediction skill is further improved against DL-Ens. This indicates that those signals interact with T2Min and can provide useful information for correcting its prediction. However, the prediction skill is not improved all over Asia, and prediction skill is worsened in some regions. This indicates that the signals used for postprocessing have regional dependency, whereas the usage of the right signal can improve the prediction for a specific region and the wrong signal may worsen the prediction in the DL models. It is interesting to see that the regions with significant improvement are significantly affected by jet and MJO documented by previous studies. For example, the 2-m air temperature over central Asia is significantly affected by MJO<sup>18</sup>. Western China and the eastern Tibetan Plateau are significantly influenced by the subtropical jet stream<sup>73</sup>. So far, less study has explored the influence of MJO on the temperature variation over eastern Russia. Except for jet and MJO, there are many other quasi-periodic signals in the atmosphere, land, and ocean, which can have significant effects on specific regions and variables. Therefore, specific DL models can be designed for specific regions with the quasi-periodic signals that affect the climate in those regions for any operational model. Although there are still some unsolved issues, it is generally clear that jet and MJO have significant influence on the air temperature over those regions, except for REG4. The present study didn't focus on a specific physical phenomenon, and thus it is difficulty to provide specific physical explanations. Though the improvement of the prediction skill seems to be specific in the DL model, the mechanism can be very complicated, which can not be explained by a single work. On the other hand, for the

DL model, the mechanism behind is a worldwide challenge for DL applied in physical world. As the technology improved fastly, the DL structure will be changed, and the mechanism for an old DL structure may not suitable for a new one.

## Conclusions

The quasi-periodic signals in the earth system are “treasures” for S2S climate prediction. The quasi-periodic features of the jet streams are explored from the daily zonal wind at 200 hPa of ERA5. Significant periodic power is identified in 10–40 days for the concurrent variation of the subtropical and polar jet streams over Asia, which includes the concurrent variation of both intensity and location. Compared to the jet streams, MJO is the dominant S2S variability in the tropics and also has quasi-periodic nature.

Like many previous studies, the results of DL-Ens show that the DL approach can find the best weights for the ensemble members to significantly improve the prediction skill than the direct mean of ensemble members. Furthermore, using reforecasts of the S2S forecast from the BCC operational model of CMA, the DL models are constructed applying quasi-periodic signals predicted by the BCC model to postprocess the T2Min prediction during boreal winter over Asia. The DL models include LSTM layers that considered the sequence dependencies in the time series, and thus the quasi-periodic signal can be passed to the future period to improve the prediction of T2Min during 11–20 days. Moreover, the advantage of the well-predicted MJO can be privileged for the improvement of the prediction of T2Min. When the jet is considered in the DL models, the prediction skill over western China and the eastern Tibetan Plateau is significantly improved compared to DL-Ens. When MJO is considered in the DL models, the prediction skill over central Asia and eastern Russia is significantly improved compared to DL-Ens. The RMSE is generally decreased by the DL models over entire Asia. This study suggests the further application of the DL approach in the postprocessing of S2S forecasts considering the usage of quasi-periodic signals as input features. Due to the regional dependency of the DL models, especially its input features, the method presented in this study can be further applied in specific regions with specific input features for the improvement of S2S prediction of an operational model or multi-models.

So far, the S2S prediction of severe cold weathers in winter is still a challenge for operational centers. Applying DL methods during the postprocessing of dynamical model outputs, to a certain extent, can improve the T2Min prediction. In other words, there is a great potential for the application of DL methods to predict extreme cold weathers. However, there are still unsolved issues for future studies. For example, while the T2Min prediction is improved, can we capture the extreme cold events? Moreover, this study only considered outputs of a single dynamical model, but the DL method is a data-driven approach. If outputs of multi-models are used, can we further improve the prediction when the quasi-periodic signals are considered? Are there other quasi-periodic signals that can benefit S2S prediction? Finally, the present study proves that quasi-periodic signals in the atmosphere do benefit the S2S operational prediction with the help of deep learning.

## Methods

**Data.** The hourly 2-m air temperature of ERA5 reanalysis<sup>74</sup> with the horizontal resolution of  $0.25^\circ \times 0.25^\circ$  during 2004–2019 is provided by the European Center for Medium-Range Weather Forecasts (ECMWF). The daily T2Min is obtained through identifying the minimum value of the hourly 2-m air temperature during 24 h. The hourly zonal wind ( $U_{\text{ERA}}$ ) at 200 hPa of ERA5 with a horizontal resolution of  $0.25^\circ \times 0.25^\circ$  during 1979–2019 is averaged into daily data for analysis. Both T2Min and  $U_{\text{ERA}}$  of  $0.25^\circ \times 0.25^\circ$  are binned averaged into the horizontal resolution of  $1.5^\circ \times 1.5^\circ$ , which is for the comparison with the S2S products of the BCC model. The ERA5 data can be downloaded on the website of <https://cds.climate.copernicus.eu/>.

Besides the ERA5 reanalysis, the daily zonal winds of the US National Center for Atmospheric Research/Department of Energy reanalysis 2 (NCEP)<sup>75</sup> at 850 and 200 hPa during 1979–2019 are with a horizontal resolution of  $2.5^\circ \times 2.5^\circ$ . The daily interpolated outgoing longwave radiation (OLR) of the US National Oceanic and Atmospheric Administration during 1979–2019 has a horizontal resolution of  $2.5^\circ \times 2.5^\circ$ . Both the NCEP zonal winds ( $U_{\text{NCEP2}}$ ) and OLR data are obtained on the website of <https://psl.noaa.gov/data/gridded/index.html>.

The S2S reforecasts of the BCC model during 2004–2019, including T2Min, OLR, and zonal winds at 850 and 200 hPa, are provided by CMA. The BCC model is a fully-coupled climate system for S2S operational prediction, which is running with fixed initial dates (twice a week) in a year. Each running is integrated for 60 days with four ensemble members, which are initialized at 18, 12, and 06 UTC of the day before the first forecast day and 00 UTC of the first forecast day. The atmospheric module of the BCC model has a horizontal resolution of approximately  $45 \times 45$  km (T266) and 56 sigma-pressure hybrid vertical levels. The outputs of the reforecast are interpolated into the horizontal resolution of  $1.5^\circ \times 1.5^\circ$  and pressure levels to satisfy the requirements of the S2S project (<http://www.s2sprediction.net>). Details about the BCC model and its S2S outputs can be referred to<sup>76</sup>. The first-30-day reforecasts initialized during boreal winter (December–February) are analyzed in the present study. It is noted that on the website of the S2S project, the BCC model provides the reforecast only during 2004–2018.

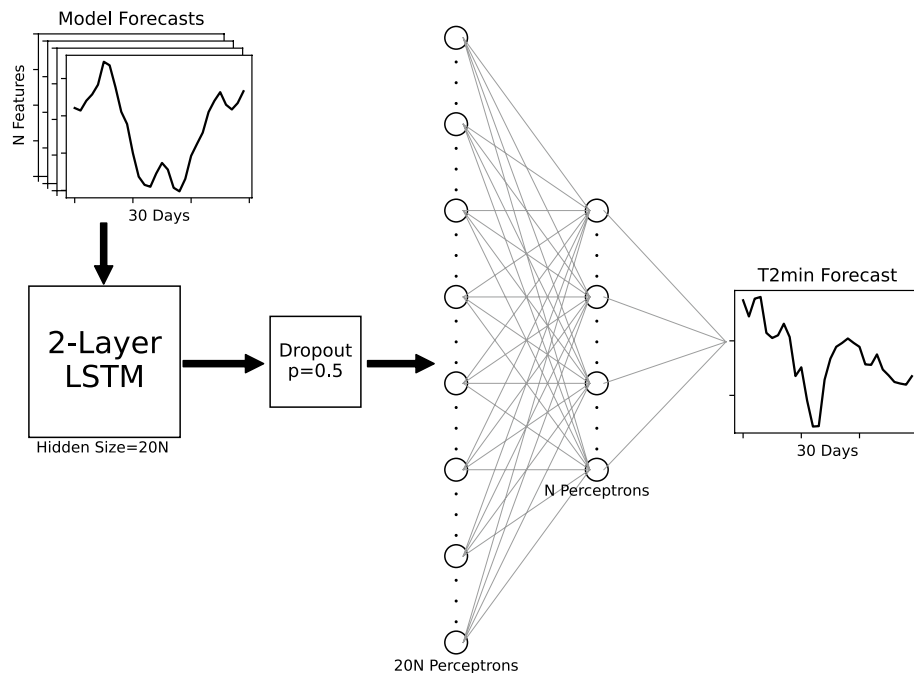
**MJO index.** The index of MJO proposed by<sup>70</sup> (WH04) for monitoring MJO is provided by the Bureau of Meteorology of Australia (<http://www.bom.gov.au/climate/mjo/>). The WH04 index is bivariate that includes two time series (RMM1&2), which are the first two principle components (PCs) of the Empirical Orthogonal Function (EOF) of OLR and zonal winds at 850 and 200 hPa of NCEP reanalysis 1 between  $15^\circ\text{S}$  and  $15^\circ\text{E}$ . Before conducting the EOF analysis, the seasonal cycles and mean are removed from all the data. For real-time monitoring, the first two EOFs derived from the data during 1979–2001 are saved, and then the real-time OLR and zonal winds are projected onto the EOFs to obtain the bivariate MJO indices.

Following the same recipe, the OLR and  $U_{\text{NCEP2}}$  at 850 and 200 hPa are used to obtain the first two EOFs during 1979–2001 in this study. After that, the data during 1979–2019 is projected on the two EOFs to obtain

the observed MJO indices. The correlation coefficients between the obtained indices and those directly downloaded from the website are 0.94 and 0.97 for RMM1 and RMM2, respectively, and the differences are negligible. For the S2S products of the BCC model, the prediction is firstly bin averaged into the horizontal resolution of  $2.5^\circ \times 2.5^\circ$  and then projected onto the first two EOFs mentioned above to obtain predicted MJO indices for the four ensemble members. The bivariate correlation of MJO indices between the ensemble prediction and observations is greater than 0.5 during the first 23 forecast days in boreal winter. This indicates that the lead time of MJO prediction with useful skill is about 23 days in the BCC model during boreal winter, which is consistent with the results of previous studies<sup>20,21</sup>.

**Jet index.** In addition to MJO, the jet streams over Asia, including subtropical and polar jet streams<sup>77</sup>, also have quasi-periodic nature on S2S time scales. In order to present their quasi-periodic feature on S2S time scales, indices are needed to represent their activities over Asia on S2S time scales. Referring to the idea of the MJO index of WH04, EOF analysis is conducted on the daily  $U_{ERA}$  with the resolution of  $1.5^\circ \times 1.5^\circ$  at 200 hPa over the region of  $49.5\text{--}148.5^\circ\text{E}$  and  $19.5\text{--}79.5^\circ\text{N}$ , which is the major region for the jet stream activities over Asia. The seasonal cycle is firstly removed from the data, and then the first two EOFs of 200-hPa  $U_{ERA}$  during the winters of 1979–2008 are obtained. After that, the data during 1979–2019 is projected on the first two EOFs to obtain bivariate indices (JET1&2) representing the activities of jet streams over Asia. For the S2S product of the BCC model, the zonal wind at 200 hPa is projected onto the first two EOFs mentioned above to obtain the prediction of JET1&2 during the winters of 2004–2019 for the four ensemble members. The bivariate correlation coefficient is calculated for JET1&2 between the model and ERA5, and the useful prediction skill for the bivariate indices of the jet has a lead time of about 10 days in the BCC model.

**Deep learning.** Using the first-30-day forecasts of T2Min, RMM1&2, and JET1&2 at a grid cell in the BCC model as input features of a DL model (Fig. 8), the prediction of T2Min is postprocessed against the T2Min of ERA5. In Fig. 8, the DL model has an input layer with the BCC model forecasts as input features (number of N features). The features are inputs of two long short-term memory (LSTM) layers with 20N as the size of the hidden features. LSTM is recurrent neural networks designed for remembering signals for a long time period, which means it is capable of learning long-term dependencies<sup>78</sup>. Thus, LSTM is suitable for processing speech recognition and the time series with dependencies<sup>79</sup>, for example quasi-periodic signals. In the present study, the LSTM models have two hidden layers that are bidirectional. To avoid overfitting the LSTM networks, a dropout layer with a dropout rate of 0.5 is set after the LSTM layers<sup>80</sup>. Following the dropout layer (Fig. 8), there



**Figure 8.** Structure of the DL models. It has an input layer (top left) with 30-day forecasts from the BCC model as input features with the number of N, including ensemble members of the prediction for T2Min, JET1&2, and MJO1&2. Following the arrows, the N features are passed to two LSTM layers with a hidden size of 20N and then a dropout layer that randomly sets elements to zero to avoid overfitting with the probability of 0.5. After that, the 20N features are passed to two linear (or multi-perceptron) layers with 20N and N perceptrons (circles), respectively. The value of N is presented in Table 1. Finally, the output layer (middle right) is the postprocessed 30-day prediction of T2Min. The activation function of hyperbolic tangent (Tanh) is used in between two layers to incorporate nonlinearity.

are two linear (or multi-perceptron) layers with 20N and N perceptrons, respectively. At last, the output layer is the postprocessed T2Min of the 30-day forecast for the BCC model. In the DL model, the activation function of hyperbolic tangent (Tanh) is used to connect two layers to incorporate nonlinearity. In order to enlarge the sample size for training, the 260 forecasts of 10 winters are randomly shuffled 10 times. After each shuffle, the 260 forecasts are resized into the size of  $26 \times 10$ , and 26 is used as the batch size. After shuffled 10 times, the forecasts have the size of  $26 \times 100$ , which have a batch size of 26 and a sample size of 100. Due to the enlargement of the sample size, the epoch used in the model is 10, which yields results that are no different from 50, 100, or 200 epochs (tested at several grid cells). Thus, the small size of the epoch is used to save the computation time. Moreover, the optimization algorithm is the implement RMSprop algorithm with a learning rate of  $1e-3$ , and the learning rate of each parameter group is decaying after every 2 epochs with a decay rate of 0.1. So far, from our trial and error, it is found that either introducing additional hidden layers of LSTM or enlarging the number of perceptrons in linear layers does not improve the predictions as additional DL-model complexity may increase the potential of overfitting.

Using the same structure (Fig. 8), four DL models with different input features are trained and used to postprocess 30-day prediction of T2Min (Table 1). For the S2S products of the BCC model, the predictions of four ensemble members are provided. In Table 1, the model DL-Ens uses the prediction of T2Min from the four ensemble members as the input features. This is synonymous with finding the best weights for the members to obtain the ensemble mean. DL-Jet and DL-MJO have the same input features of 12. Each model includes the predictions of both the T2Min and one kind of the bivariate indices from the four ensemble members, and thus effects of each kind of the indices on the results of postprocessing can be considered. The model DL-All uses T2Min, JET1&2, and RMM1&2 from the four ensemble members as input features, which is of the number 20.

During the winters of 2004–2019, the forecasts initiated in January and February of 2004 and December of 2019 are discarded. Therefore, there are 15 winters with 390 forecasts for analysis. Instead of randomly choosing the samples for training and to mimic operational conditions as closely as possible, the model forecast of the first 10 winters with the number of 260 forecasts is used for training the DL models to fit ERA5 T2Min, and the last five winters with the number of 130 forecasts are used for the DL model validation. Generally, 260 forecasts are used as training samples, and each forecast has 30 days of prediction, which means the sequence length for LSTM is 30. Furthermore, the 130 forecasts are used as testing samples to validate the performance of the DL models. The correlation coefficient (R) and root mean square error (RMSE) between the T2Min from the DL models and ERA5 are calculated for validation.

**Significant test.** During the validation for the DL forecast of the last five winters, the significance of the improvement is investigated. In order to check whether the result of DL-Ens is significantly better than the direct mean of the four ensemble members on a forecast day or during a forecast period, the following steps are conducted. (1) The R and RMSE of DL-Ens are calculated (denoted as  $R_{DL-Ens}$  and  $RMSE_{DL-Ens}$ , respectively) for the five winters, as well as those for the direct mean of the four ensemble members (denoted as  $R_{Ens}$  and  $RMSE_{Ens}$ , respectively). The differences are calculated between  $R_{DL-Ens}$  and  $R_{Ens}$  ( $\Delta R_{DL-Ens}$ ), as well as those between  $RMSE_{DL-Ens}$  and  $RMSE_{Ens}$  ( $\Delta RMSE_{DL-Ens}$ ). (2) Due to the DL-Ens is equal to seek the best weights for the four ensemble members, four float values (seven decimal places) are randomly chosen between 0.0 and 1.0 and divided by their sum as the weights for the four ensemble members. This procedure is conducted 1000 times to obtain 1000 predictions through summing the four ensemble members with random weights. (3) The R and RMSE of the 1000 predictions are calculated, denoted as  $R_i$  and  $RMSE_i$  ( $1 \leq i \leq 1000$ ). Meanwhile, the differences are calculated between  $R_i$  and  $R_{Ens}$  ( $\Delta R_i$ ), as well as those between  $RMSE_i$  and  $RMSE_{Ens}$  ( $\Delta RMSE_i$ ). (4) The percentages of the  $\Delta R_i \geq \Delta R_{DL-Ens} > 0$  and  $\Delta RMSE_i \leq \Delta RMSE_{DL-Ens} < 0$  are identified, denoted as  $p_r$  and  $p_{rmse}$ , respectively. If the  $p$  value is smaller than 5%, it means that DL-Ens does not improve the forecast by chance, which indicates DL-Ens significantly improves the forecast at the 5% level. On the contrary, the percentages of the  $\Delta R_i \leq \Delta R_{DL-Ens} < 0$  and  $\Delta RMSE_i \geq \Delta RMSE_{DL-Ens} > 0$  can be also identified. If the  $p$  value is smaller than 5%, it means that DL-Ens does not worsen the forecast by chance, which indicates DL-Ens significantly worsen the forecast at the 5% level.

For the DL-Jet, DL-MJO, and DL-All, whether including MJO or jet signals can significantly improve the prediction of DL-Ens is examined. The procedure is similar to that for the significant test of DL-Ens, except for some differences. The time sequences of the indices of MJO and jet are randomly shuffled 1000 times, and then the DL models are trained by those shuffled time sequences for obtaining  $R_i$  and  $RMSE_i$ , and  $\Delta R_i$  and  $\Delta RMSE_i$  are calculated against  $R_{DL-Ens}$  and  $RMSE_{DL-Ens}$ , respectively. After that, the  $p$  values for the significant test are obtained. During this procedure, we can ensure that whether those indices can significantly improve the prediction. Due to this procedure trains the model 1000 times at each grid cell, large consumption of computation resources is needed.

### Data availability

All the data used can be downloaded freely from the website. ERA5 reanalysis is on the website of <https://cds.climate.copernicus.eu/>. NCEP reanalysis and OLR are downloaded on the website of <https://psl.noaa.gov/data/gridded/index.html>. S2S products of the BCC model can be found on the website of <http://www.s2sprediction.net>. The WH04 MJO index is on <http://www.bom.gov.au/climate/mjo/>.

Received: 8 September 2022; Accepted: 10 March 2023

Published online: 13 March 2023

## References

- Vitart, F. *et al.* The subseasonal to seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.* **98**, 163–173 (2017).
- White, C. J. *et al.* Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorol. Appl.* **24**, 315–325 (2017).
- White, C. J. *et al.* Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Am. Meteor. Soc.* **103**(6), E1448–E1472 (2022).
- Madden, R. A. & Julian, P. R. Observation of the 40–50-day tropical oscillation—A review. *Mon. Wea. Rev.* **122**, 814–837 (1994).
- Zhang, C. Madden–Julian oscillation. *Rev. Geophys.* **43**, RG2003 (2005).
- Lin, H. The Madden–Julian oscillation. *Atmos. Ocean* <https://doi.org/10.1080/07055900.2022.2072267> (2022).
- Vitart, F. Madden–Julian oscillation prediction and teleconnections in the S2S database. *Q. J. R. Meteorol. Soc.* **143**, 2210–2220 (2017).
- Xiang, B. *et al.* S2S prediction in GFDL SPEAR: MJO diversity and teleconnections. *Bull. Am. Meteor. Soc.* **103**(2), E463–E484 (2022).
- Zhou, Y. *et al.* Effects of the Madden–Julian oscillation on 2-m air temperature prediction over china during boreal winter in the S2S database. *Clim. Dyn.* **52**, 6671–6689 (2019).
- Zhou, Y. & Wang, Y. Influence of the Madden–Julian oscillation on the arctic oscillation prediction in S2S operational models. *Front. Earth Sci.* **9**, 787680 (2021).
- Specq, D. & Batté, L. Do subseasonal forecasts take advantage of Madden–Julian oscillation windows of opportunity?. *Atmos. Sci. Lett.* **23**, e1078 (2022).
- Lorenz, E. N. Climatic predictability. The physical Basis of Climate and Climate Modeling, GARP Publication Series, Vol. 16, World Meteorological Organization, 132–136 (1975).
- von Neumann, J. Some remarks on the problem of forecasting climate fluctuations. Paper presented at Dynamics of Climate: The Proceedings of a Conference on the Application of Numerical Integration Techniques to the Problem of the General Circulation. Pergamon Press, Oxford, U.K. (published 1960).
- Waliser, D. E., Jones, C., Schemm, J.-K.E. & Graham, N. E. A statistical extended-range tropical forecast model based on the slow evolution of the Madden–Julian oscillation. *J. Climate* **12**, 1918–1939 (1999).
- Pegion, K. & Sardeshmukh, P. D. Prospects for improving subseasonal predictions. *Mon. Wea. Rev.* **139**, 3648–3666 (2011).
- Stan, C. *et al.* Advances in the prediction of MJO teleconnections in the S2S forecast systems. *Bull. Amer. Meteor. Soc.* <https://doi.org/10.1175/BAMS-D-21-0130.1> (2022).
- Zhou, Y., Thompson, K. R. & Lu, Y. Mapping the relationship between Northern hemisphere winter surface air temperature and the Madden–Julian oscillation. *Mon. Wea. Rev.* **139**, 2439–2454 (2011).
- Zhou, Y. *et al.* On the relationship between the Madden–Julian Oscillation and 2 m air temperature over central Asia in boreal winter. *J. Geophys. Res. Atmos.* **121**, 13250–13272 (2016).
- Tseng, K.-C., Barnes, E. A. & Maloney, E. The importance of past MJO activity in determining the future state of the midlatitude circulation. *J. Clim.* **33**, 2131–2147 (2020).
- Wu, J. *et al.* Effects of moisture initialization on MJO and its teleconnection prediction in BCC subseasonal coupled model. *J. Geophys. Res. Atmos.* **125**, e2019JD031537 (2020).
- Wu, J. & Jin, F.-F. Improving the MJO forecast of S2S operation models by correcting their biases in linear dynamics. *Geophys. Res. Lett.* **48**, 091930 (2021).
- Cohen, J. *et al.* S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecast. *WIREs Clim. Change* **10**, e567 (2019).
- Kim, H., Ham, Y. G., Joo, Y. S. & Son, S. W. Deep learning for bias correction of MJO prediction. *Nat. Commun.* **12**, 3087 (2021).
- Weyn, J. A., Durran, D. R., Caruana, R. & Cresswell-Clay, N. Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *J. Adv. Model. Earth Syst.* **13**, 002502 (2021).
- Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
- Rencher, A. C. & Schaalje, C. B. *Linear Models in Statistics* 2nd edn, 688 (Jonh Wiley & Sons, New York, 2008).
- Wilks, D. *Statistical Methods in the Atmospheric Sciences* 4th edn, 818 (Elsevier Inc., New York, 2019).
- Mu, R. & Zeng, X. A review of deep learning research. *KSII Trans. Internet Inf. Syst.* **13**, 1738–1764 (2019).
- Ten Perkel, J. M. Computer codes that transformed science. *Nature* **589**, 344–348 (2021).
- Bonavita, M. *et al.* Machine learning for earth system observation and prediction. *Amer. Meteor. Soc. Meet. Summary Bull.* <https://doi.org/10.1175/BAMS-D-20-0307.1> (2020).
- Sit, M. *et al.* A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.* **82**, 2635–2670 (2020).
- Yuan, Q. *et al.* Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **241**, 111716 (2020).
- Shakya, S., Kumar, S. & Goswami, M. Deep learning algorithm for satellite imaging based cyclone detection. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* **13**, 827–839 (2020).
- Chattopadhyay, A., Hassanzadeh, P. & Pasha, S. Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Sci. Rep.* **10**, 1317 (2020).
- Li, W., Gao, X., Hao, Z. & Sun, R. Using deep learning for precipitation forecasting based on spatio-temporal information: A case study. *Clim. Dyn.* **58**, 443–457 (2022).
- Stringari, C. E., Guimaraes, P. V., Filipot, J.-F., Leckler, F. & Duarte, R. Deep neural networks for active wave breaking classification. *Sci. Rep.* **11**, 3604 (2021).
- Chen, H., Chandrasekar, V., Tan, H. & Cifelli, R. Rainfall estimation from ground radar and TRMM precipitation radar using hybrid deep neural networks. *Geophys. Res. Lett.* **46**, 10669–10678 (2019).
- Wimmers, A. & Velden, C. Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Wea. Rev.* **147**, 2261–2282 (2019).
- Dawood, M., Asif, A. & Minhas, F. Deep-PHURIE: Deep learning based hurricane intensity estimation from infrared satellite imagery. *Neural Comput. Appl.* **32**, 9009–9017 (2020).
- Yen, M.-H., Liu, D.-W., Hsin, Y.-C., Lin, C.-E. & Chen, C.-C. Application of the deep learning for the prediction of rainfall in southern Taiwan. *Sci. Rep.* **9**, 12774 (2019).
- Miao, K. *et al.* Multimodal semisupervised deep graph learning for automatic precipitation nowcasting. *Math. Probl. Eng.* **2020**, 4018042 (2020).
- Ravuri, S. *et al.* Skilful precipitation now casting using deep generative models of radar. *Nature* **597**, 672–677 (2021).
- Tran, T., Bateni, S., Ki, S. & Vosoughifar, H. A review of neural networks for air temperature forecasting. *Water* **13**, 1294 (2021).
- Scher, S. Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophys. Res. Lett.* **45**, 12616–12622 (2018).
- Weyn, J., Durran, D. & Caruana, R. Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J. Adv. Model. Earth Syst.* **11**, 2680–2693 (2019).
- Schultz, M. G. *et al.* Can deep learning beat numerical weather prediction?. *Phil. Trans. R. Soc. A* **379**, 20200097 (2021).

47. Kashinath, K. *et al.* Physics-informed machine learning: Case studies for weather and climate modelling. *Phil. Trans. R. Soc.* **A379**, 20200093 (2021).
48. Fablet, R. *et al.* Learning variational data assimilation models and solvers. *J. Adv. Model. Earth Syst.* **13**, e2021MS002572 (2021).
49. Rasp, S. & Lerch, S. Neural networks for postprocessing ensemble weather forecasts. *Mon. Wea. Rev.* **146**, 3885–3900 (2018).
50. Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. & Yacalis, G. Could machine learning break the convection parameterization deadlock?. *Geophys. Res. Lett.* **45**, 5742–5751 (2018).
51. Bonavita, M. & Laloyaux, P. Machine learning for model error inference and correction. *J. Adv. Model. Earth Syst.* **12**, e2020MS002232 (2020).
52. Rasp, S., Pritchard, M. S. & Gentine, P. Deep learning to represent subgrid processes in climate models. *PNAS* **115**, 9684–9689 (2018).
53. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–584 (2019).
54. Tang, Y. & Duan, A. Using deep learning to predict the East Asian summer monsoon. *Environ. Res. Lett.* **16**, 124006 (2021).
55. Saha, M. & Nanjundiah, R. S. Prediction of the ENSO and EQUINOO indices during June–September using a deep learning method. *Meteorol. Appl.* **27**, e1826. <https://doi.org/10.1002/met.1826> (2020).
56. Zhang, S. *et al.* Linkage of extreme temperature change with atmospheric and locally anthropogenic factors in China mainland. *Atmos. Res.* **277**, 106307 (2022).
57. Yi, X. *et al.* Multi-model ensemble projections of winter extreme temperature events on the Chinese mainland. *Int. J. Environ. Res. Public Health* **19**, 5902 (2022).
58. Kuang, X., Zhang, Y., Huang, D. & Huang, Y. Regionality of record-breaking low temperature events in China and its associated circulation. *Clim. Dyn.* **46**, 1719–1731 (2016).
59. Strong, C. & Davis, R. E. Winter jet stream trends over the Northern Hemisphere. *Q. J. R. Meteorol. Soc.* **133**, 2109–2115 (2007).
60. Ma, X. & Zhang, Y. Interannual variability of the North Pacific winter storm track and its relationship with extratropical atmospheric circulation. *Clim. Dyn.* **51**, 3685–3698 (2018).
61. Hannachi, A. & Iqbal, W. On the nonlinearity of winter northern hemisphere atmospheric variability. *J. Atmos. Sci.* **76**, 333–356 (2019).
62. Bushra, N. & Rohli, R. V. Relationship between atmospheric teleconnections and the Northern Hemisphere’s circumpolar vortex. *Earth Space Sci.* **8**, e2021EA001802 (2021).
63. Hallam, S., Josey, S. A., McCarthy, G. D. & Hirschi, J. J. M. A regional (land–ocean) comparison of the seasonal to decadal variability of the Northern Hemisphere jet stream 1871–2011. *Clim. Dyn.* **59**, 1897–1918 (2022).
64. Zhang, Y. C., Wang, D. Q. & Ren, X. J. Seasonal variation of the meridional wind in the temperate jet stream and its relationship to the Asian monsoon. *Acta Meteor. Sin.* **22**, 446–454 (2008).
65. Hu, K., Huang, G., Wu, R. & Wang, L. Structure and dynamics of a wave train along the wintertime Asian jet and its impact on East Asian climate. *Clim. Dyn.* **51**, 4123–4137 (2018).
66. Dong, X., Zhao, P. & Ren, H.-L. Climatic factors contributing to interannual and interdecadal variations in the meridional displacement of the East Asian jet stream in boreal winter. *Atmos. Res.* **264**, 105864 (2021).
67. Cohen, J. *et al.* Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.* **7**, 627–637 (2014).
68. Overland, J. E. *et al.* How do intermittency and simultaneous processes obfuscate the Arctic influence on midlatitude winter extreme weather events?. *Environ. Res. Lett.* **16**, 043002 (2021).
69. Green, M. R. & Furtado, J. C. Evaluating the joint influence of the Madden-Julian oscillation and the stratospheric polar vortex on weather patterns in the Northern hemisphere. *J. Geophys. Res.: Atmos.* **124**, 11693–11709 (2019).
70. Wheeler, M. C. & Hendon, H. H. An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.* **132**, 1917–1932 (2004).
71. Ha, K.-J. *et al.* Variability in the East Asian monsoon: A review. *Meteorol. Appl.* **19**, 200–215 (2012).
72. Wu, S. & Sun, J. Variability in zonal location of winter East Asian jet stream. *Int. J. Climatol.* **37**, 3753–3766 (2017).
73. Schiemann, R. *et al.* Seasonality and interannual variability of the westerly jet in the Tibetan Plateau region. *J. Climate* **22**, 2940–2957 (2009).
74. Hans, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
75. Kanamitsu, M. *et al.* NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.* **83**, 1631–1643 (2002).
76. Liu, X. *et al.* Development of coupled data assimilation with the BCC climate system model: Highlighting the role of sea-ice assimilation for global analysis. *J. Adv. Model. Earth Syst. (JAMES)* **13**(4), e2020MS002368 (2021).
77. Molnos, S. *et al.* A network-based detection scheme for the jet stream core. *Earth Syst. Dynam.* **8**, 75–89 (2017).
78. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
79. Sak, H., *et al.* Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. [arXiv:1402.1128](https://arxiv.org/abs/1402.1128) (2014). <https://doi.org/10.48550/arXiv.1402.1128>
80. Hinton, G., *et al.* Improving neural networks by preventing co-adaptation of feature detectors. [arXiv:1207.0580](https://arxiv.org/abs/1207.0580) (2012). <https://doi.org/10.48550/arXiv.1207.0580>

## Acknowledgements

We thank the support from the National Natural Science Foundation of China (Grant No. 41930969 and 42175030).

## Author contributions

Y. Z. contributed the basic idea, coding, and writing. Q. Z. contributed to data processing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Y.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023