

RESEARCH LETTER

Open Access



Application of a weighted ensemble forecasting method based on online learning in subseasonal forecast in the South China

Fei Xin^{1,2}, Yichen Shen⁴ and Chuhan Lu^{3*}

Abstract

Under the proposal of “seamless forecasting”, it has become a key problem for meteorologists to improve the skills of subseasonal forecasts. Since the launch of the subseasonal-to-seasonal (S2S) plan by WMO, the precision of model predictions has been further developed. However, when we are focusing on the practical applications of models in the South China (SC) in recent years, we found that large disagreements appear between forecast members. Some of the members predicted well in this area, while others are not satisfactory. To improve the accuracy of subseasonal forecast in the SC, new methods making full use of different forecast models must be proposed. In this passage, we introduced a weighted ensemble forecasting method based on online learning (OL) to overcome this difficulty. As the state-of-the-art forecast models in the world, three models from China Meteorological Administration, European Centre for Medium-Range Weather Forecasts and National Centers for Environmental Prediction provided by the S2S prediction dataset are used as ensemble members, and an ensemble weight is trained through the aforementioned OL model for the predictions of temperature and precipitation in subseasonal timescale in the SC. The results show that the forecast results produced under the OL method are better than the original model predictions. Compared with the three model ensemble results, the weighted ensemble model has a good ability in depicting the temperature and precipitation in the SC. Furthermore, we also compared this strategy against the climatology predictions and found out that the weighted ensemble model is superior in 10–30 days. Thus, the weighted ensemble method trained thorough OL may shed light on improving the skill of subseasonal forecasts.

Keywords Online learning, Subseasonal forecast, Yangtze River Delta, Weighted ensemble forecast

Introduction

The extended-range forecast with a lead time of 10–30 days is the gap between weather (within 10 days) and climate (beyond 30 days) predictions (Xie et al. 2023) and is a crucial part in constructing the seamless forecasting system but also a vulnerable spot (Hoskins 2013). To accelerate the progress in subseasonal-to-seasonal (S2S) forecast and fill in this gap the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP) jointly launched the S2S Prediction Project to improve the prediction ability and understand the predictability source (Vitart et al. 2017). Relying on this program, model prediction results from different countries are released regularly, thus making

*Correspondence:

Chuhan Lu
luchuhan@nuist.edu.cn

¹ Key Laboratory of Cities Mitigation and Adaptation to Climate Change in Shanghai (CMACC), China Meteorological Administration, Shanghai Climate Center, Shanghai 200030, China

², Shanghai Climate Center, Shanghai 200030, China

³ Key Laboratory of Ecosystem Carbon Source and Sink, China Meteorological Administration (ECSS-CMA), Wuxi University, Wuxi 214063, China

⁴ School of Atmospheric Science, Nanjing University of Information Science and Technology, Nanjing 210044, China



© The Author(s) 2024. **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 localized interpretation and application of ensemble model prediction possible.

However, according to previous researches, when applied to a certain area, these model predictions inevitably exhibit biases in different seasons. For example, Climate Forecast System version 2 (CFSv2) predictions show a warm bias for the climatological mean SAT over the Yangtze River valley (Tang et al. 2020), and it is also deficient in depicting the winter SAT in most part of mainland China (Zou et al. 2022). Facing such problems, one effective way is to make ensemble predictions, which is widely adopted by meteorological bureaus around the world. It is obvious that some models will perform better than others and deserve more weight, so there lies a question that what weight is appropriate? Also, the S2S dataset is only providing data at a relatively low resolution of $1.5^{\circ} \times 1.5^{\circ}$, which is too coarse for practical application. Can we somehow increase its resolution to meet the operation requirements?

In the recent years, machine learning methods is increasingly used in meteorological applications, among which artificial intelligence (AI) and online learning (OL) are two cutting-edge technologies. In weather forecasting, there are already neural network models being proposed for system identification (Lu et al. 2020), nowcasting (Zhang et al. 2023), statistical downscaling (Baño-Medina et al. 2020) and so on. As a classic structure in neural network, UNet structure is first created for biomedical image segmentation (Ronneberger et al. 2015), but later introduced into meteorological field for downscaling (Sha et al. 2020a, 2020b), owing its ability to reconstruct the high-resolution data to the hierarchical decoders and skip connections. Such applications give us a hint that neural network model may also help in dealing with model predictions for high-resolution forecasts.

OL, a novel training strategy, is a sequential decision-making paradigm that is widely used in webpage predictions and other Internet applications (Orabona 2019). It can dynamically and real-timely adjust the model according to data feedbacks and thus reflect the changes in input data stream and improved the accuracy (Graepel et al. 2010, McMahan et al. 2013). In other words, unlike traditional training methods like neural networks that generates static models according to a static database, online learning generates dynamical models using dynamical data streams. Flaspohlar et al. (2021) introduced this method into weather prediction with two new mechanisms. In their passage, Flaspohlar et al. successfully made subseasonal ensemble forecast with six models and shed light on the improvement of prediction skills for subseasonal forecasts.

In this passage, we will be focusing on the following questions. How to determine the weights for the

ensemble extended-range forecasts upon the basis of the S2S database? How to make high-resolution forecasts out of low-resolution data? Do the ensemble predictions work better than normal predictions? Based on these questions, the rest part of the paper is organized as follows. In part 2 there will be an introduction of the dataset and the method. In part 3 the results of the predictions will be shown through indices and case studies. A discussion and conclusion part will be in Part 4.

Data and methods

Modified UNet for statistical downscaling

The traditional UNet structure was modified in this study for statistical downscaling, where two deconvolution block is added at the front and back as shown in Fig. 1. This is to fit the change in size for the input and output data, as for the original structure, the input data and output data have the same size. All the activate functions have been swapped to LeakyReLU. In this study, we are focusing on the precipitation and temperature of Yangtze River Delta. The input data is the forecast within 109.5° – 126° E, 21° – 37.5° N, with a low resolution of $1.5^{\circ} \times 1.5^{\circ}$, or $12^{\circ} \times 12^{\circ}$ in grid. The output data has a high resolution of $0.0625^{\circ} \times 0.0625^{\circ}$, within 109.53125° – 126.03125° E, 21.03125° – 37.53125° N (shown in Fig. 2), or $265^{\circ} \times 265^{\circ}$ in grids.

The training set for UNet downscaling includes hourly high-resolution and low-resolution data corresponding to the model input and output mentioned above. It derives from ERA5 precipitation and 2 m temperature data in 2020–2021, including 17544 samples (Hersbach et al. 2020). This set is separated into two parts. For January 1 2020 to September 30 2021, this part of the data is used directly for training, and the rest part is for validation. Moreover, the downscaling results will be compared against common bilinear interpolation for January 1 to June 30 2022 with altogether 4344 samples.

Models and data for online learning

Three models, Beijing Climate Center Climate Prediction System version 2 for S2S (BCC-CPS-S2Sv2) model from China Meteorological Administration (CMA) (Liu et al. 2021), Integrated Forecasting System (IFS) from European Centre for Medium-Range Weather Forecasts (ECMWF) (Roberts et al. 2018) and CFSv2 model from National Centers for Environmental Prediction (NCEP) (Saha et al. 2014) are selected for the ensemble forecast, with all the historical forecasts provided by the ECMWF S2S data set. The BCC model contains 4 members, while IFS and CFSv2 have 51 and 16 members, respectively. The training set contains daily global precipitation and 2 m temperature data in 2018–2021

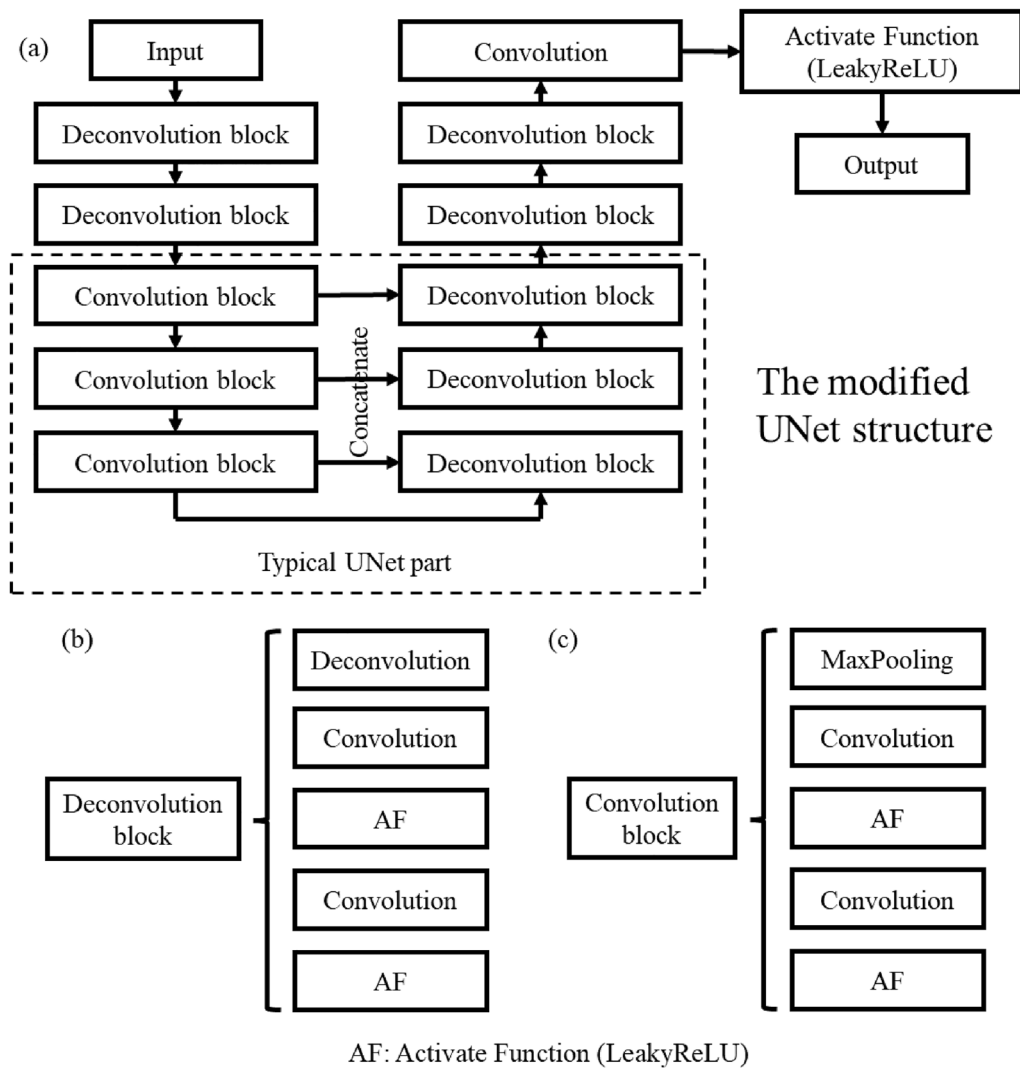


Fig. 1 (a) Structure of the modified UNet model. (b) & (c) Definition of the deconvolution block and the convolution block in (a)

for all 71 members with a resolution of $1.5^{\circ} \times 1.5^{\circ}$. The valid set contains the same data for January 2022 to June in 2023.

The historical real data in the same time range come from CLDAS data set (Shi et al. 2019) with a resolution of $0.0625^{\circ} \times 0.0625^{\circ}$. In this study, we are mainly focusing on the Yangtze River Delta region, located in the eastern part of China.

As we mentioned in Part 1, online learning is a training strategy that can dynamically and real-time adjust the model according to data feedbacks. In Flashphor's design, two new mechanisms called delay and hint are added. Delay mechanism is a simulation of the lead time. Unlike webpage predictions where we can get predictions and real data almost at the same time, weather predictions have lead times, which means

whether a prediction is accurate cannot be told until the forecast target date comes. The hint mechanism is a complement for the delay in getting real data, it allows the learner to estimate the possible loss for a prediction whose target date has not come yet. The flow chart below describes the whole process of online learning (Fig. 3).

Online learning steps

The whole process included in online learning is shown as below:

- 1) Downscale all the low-resolution ($1.5^{\circ} \times 1.5^{\circ}$) training data from S2S system to high-resolution of $0.0625^{\circ} \times 0.0625^{\circ}$ using modified UNet model.

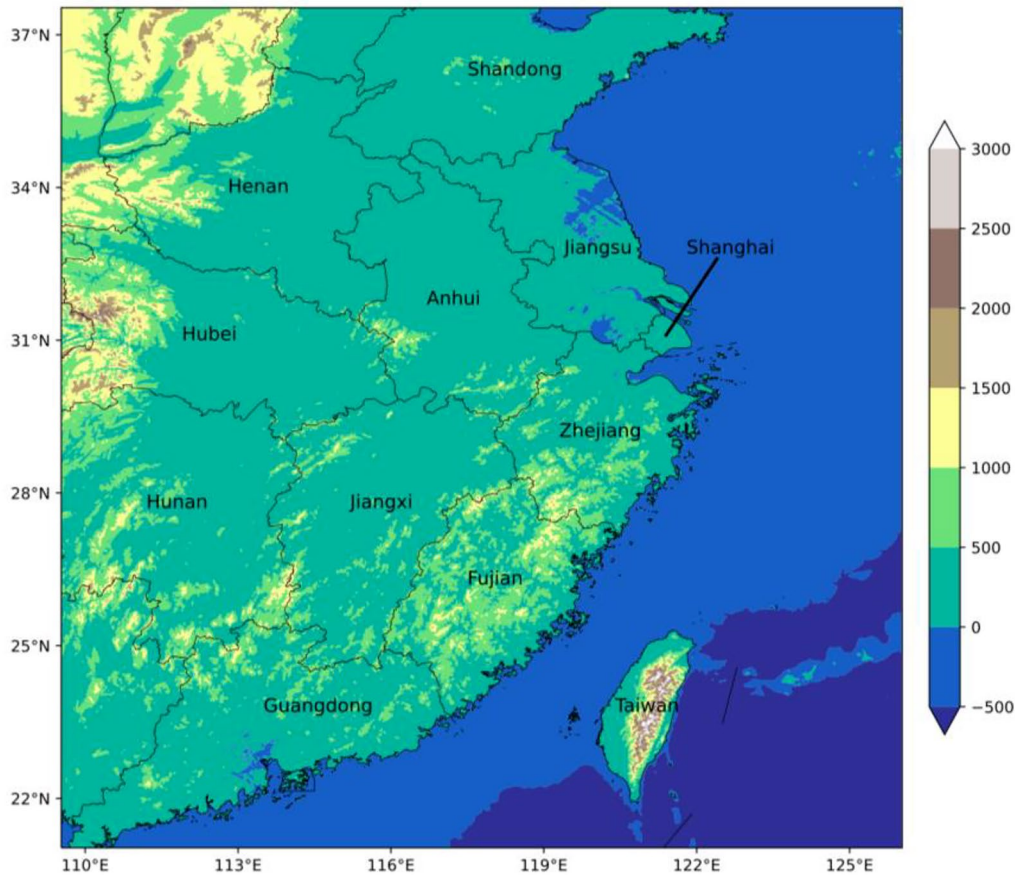


Fig. 2 Topography of the study area (shading, m), mainly covering Yangtze River Delta (109.53125–126.03125°E, 21.03125–37.53125°N)

- 2) For each lead time and each of the 71 members, calculate latitude weighted anomaly correlation coefficient (ACC) for 2018–2021. For each lead time, select all the members that surpass the average ACC into the next step.
- 3) For each lead time, for all the 12 matches of 3 hint algorithms (mean_g, recent_g, prev_g) and 4 optimize algorithms (adahedged (Orabona 2019), dorm, dormplus (Bowling et al. 2015; Zinkevich et al. 2007), dub), train 12 weight for each lead time for each match of hint algorithm and optimize algorithm.
- 4) For each lead time, use the 12 weights to make hindcasts for 2018–2021, and for each of the hindcasts, calculate latitude weighted root mean squared error (RMSE) and select the weight with least RMSE as the final result.

Evaluation methods

To evaluate the forecast results of prediction and UNet statistical downscaling, RMSE is defined as

$$RMSE = \frac{1}{N_{forecasts}} \sum_i \sqrt{\frac{N_{lat} N_{lon}}{\sum_j \sum_k L(j) (f_{i,j,k} - t_{i,j,k})^2}}$$

where f is the prediction result or the downscaling result of the model, and t is the CLDAS data for prediction or ERA5 high-resolution data for downscaling of the corresponding time. The smaller the RMSE value is, the better the prediction result of the model is. The ACC is defined as

$$ACC = \frac{\sum_{i,j,k} L(j) f'_{i,j,k} t'_{i,j,k}}{\sqrt{\sum_{i,j,k} L(j) f'^2_{i,j,k} \sum_{i,j,k} L(j) t'^2_{i,j,k}}}$$

where the symbol ' represents the difference to the climatology, $L(j)$ is the weight factor when latitude is j , and $L(j)$ is defined as

$$L(j) = \frac{\cos(lat(j))}{\frac{1}{N_{lat}} \sum_j^{N_{lat}} \cos(lat(j))}$$

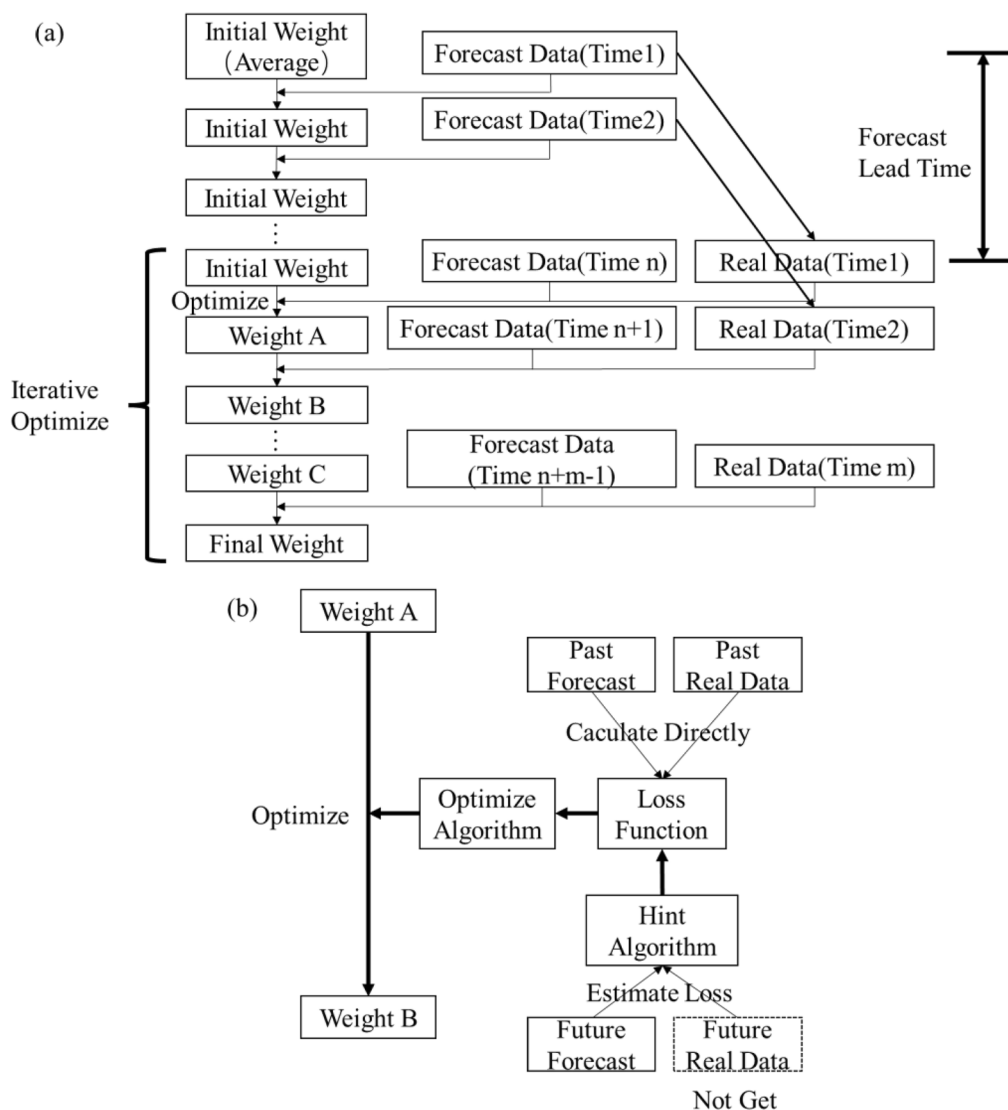


Fig. 3 a The process of online learning show in flow chart. b The optimize process in each iteration with optimize algorithm and hint algorithm

ACC can represent the similarity of two fields. The closer the absolute value of ACC is to 1, the more similar the two fields are.

Performance evaluation

Performance of the UNet statistical downscaling

The performance of the UNet downscaling is examined with ERA5 data in the first half of 2022. For the two meteorological elements we are focusing on, precipitation and 2 m temperature, 2 models are trained for each element. The calculation of latitude weighted RMSE and ACC of UNet downscaling and bilinear interpolation is shown in Table 1. We can see that except for the ACC in precipitation, all the scores UNet get are better than

Table 1 Comparison of results of UNet downscaling and bilinear interpolation in the first half of 2022

	Precipitation		2 m temperature	
	Bilinear	UNet	Bilinear	UNet
RMSE	9.45797	8.82330*	1.02371	0.89849*
ACC	0.00595	0.00581	0.99099	0.99304*

The * denotes that the difference between UNet downscaling score and the bilinear interpolation score passed the 95% significance test

traditional bilinear interpolation, proving that the modified UNet model is suitable for downscaling the low-resolution forecasts into high-resolution forecasts.

Figure 4 shows a comparison case of the two downscaling methods for both the elements. It can be seen that the high-resolution precipitation data demonstrates a belt-shaped area for precipitation above 40 mm/day (Fig. 4a), which is well depicted in UNet downscaling results in Fig. 4c. However, in the results of bilinear interpolation shown in Fig. 4b, such characteristic is not fully shown. Also, in the comparison of 2 m temperature, there is a cold tongue in the western part of Zhejiang Province (Fig. 4d), which is relatively accurately shown in UNet downscaling results in Fig. 4f together with a small cold zone below 6 °C. In the bilinear interpolation results in Fig. 4e, the shape of the cold tongue is distorted and the cold zone is missed. These two comparisons show the ability of UNet Downscaling to depict the details of the elements, which is superior to the traditional bilinear interpolation, proving that it is suitable for the downscaling processes of the S2S forecasts.

Results of the online learning ensemble forecast

The overall performance of the online learning ensemble forecast, the ERA5 climatology forecast (calculated using 1990–2019 data) and three model ensemble forecasts is compared by calculating RMSE and ACC for the

hindcasts of the valid set, as shown in Fig. 5. It can be seen that among the three model ensemble forecasts, ECMWF IFS model has the best performance, followed by NCEP CFS model and CMA model. Our online ensemble is slightly superior to IFS model during the forecast lead time of 10–30 days for both precipitation and 2 m temperature. The RMSE for precipitation is by 0.003 lower than ECMWF forecast in average and lower by 0.053 for 2 m temperature. Compared with the equal-weighted ensemble of all 71 ensembles and equal-weighted ensemble of selected members (the selection here refers to the second step in the online learning process, see part 2.3), our method still has an advantage in most lead times. The RMSE for precipitation is lower than equal-weight ensemble of selected members for 0.013, while for 2 m temperature, the advantage is 0.032. Compared with all 71 members ensemble, we also have an advantage in 2 m temperature with 0.028 smaller in RMSE for 2 m temperature prediction. Also, compared with the ECMWF ensemble forecast, the advantage for the average ACC score for the lead time of 18–26 days for both the precipitation and 2 m temperature and the average RMSE score for 2 m temperature during the same lead time period passed the 90% significance test, considering that it is a

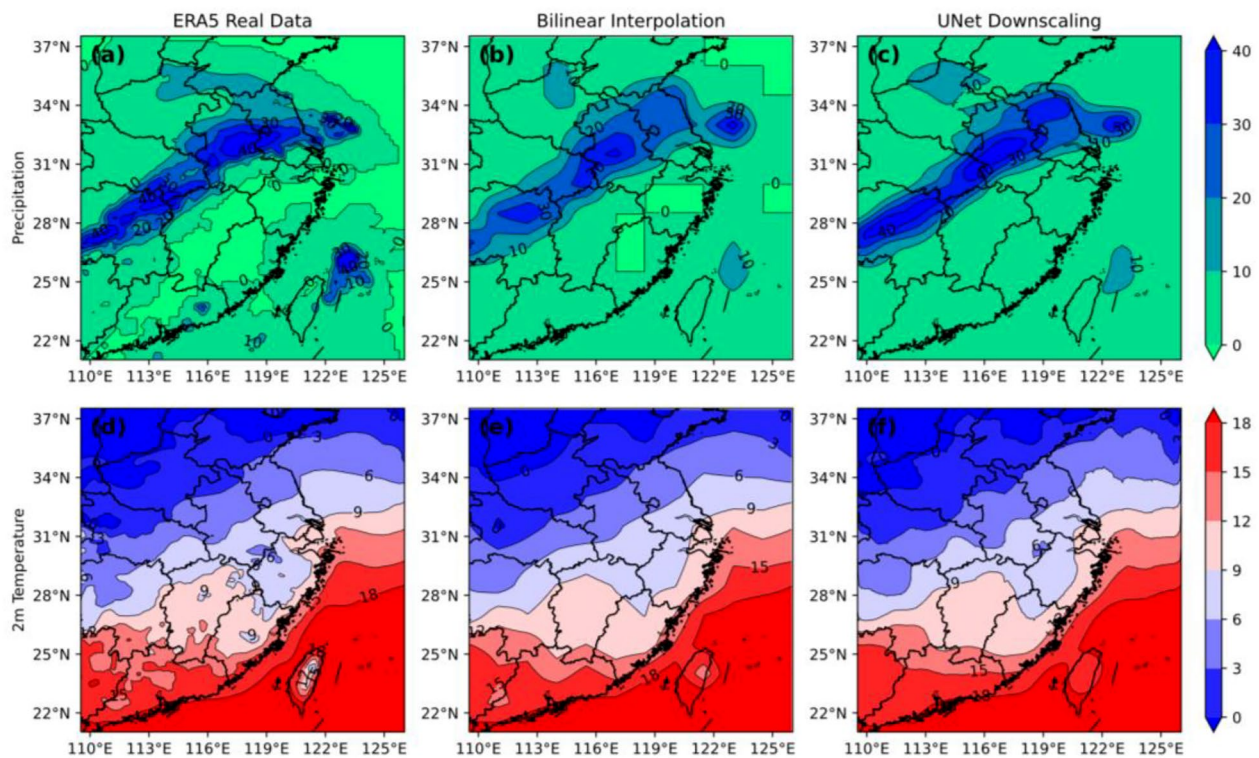


Fig. 4 a–c Comparison of the ERA5 high-resolution precipitation data and the results of bilinear interpolation and UNet downscaling (shading, mm/day). d–f Comparison of the ERA5 high-resolution 2 m temperature data and the results of bilinear interpolation and UNet downscaling (shading, °C). Both of the comparisons are done using data in 18:00:00 January 4 2022

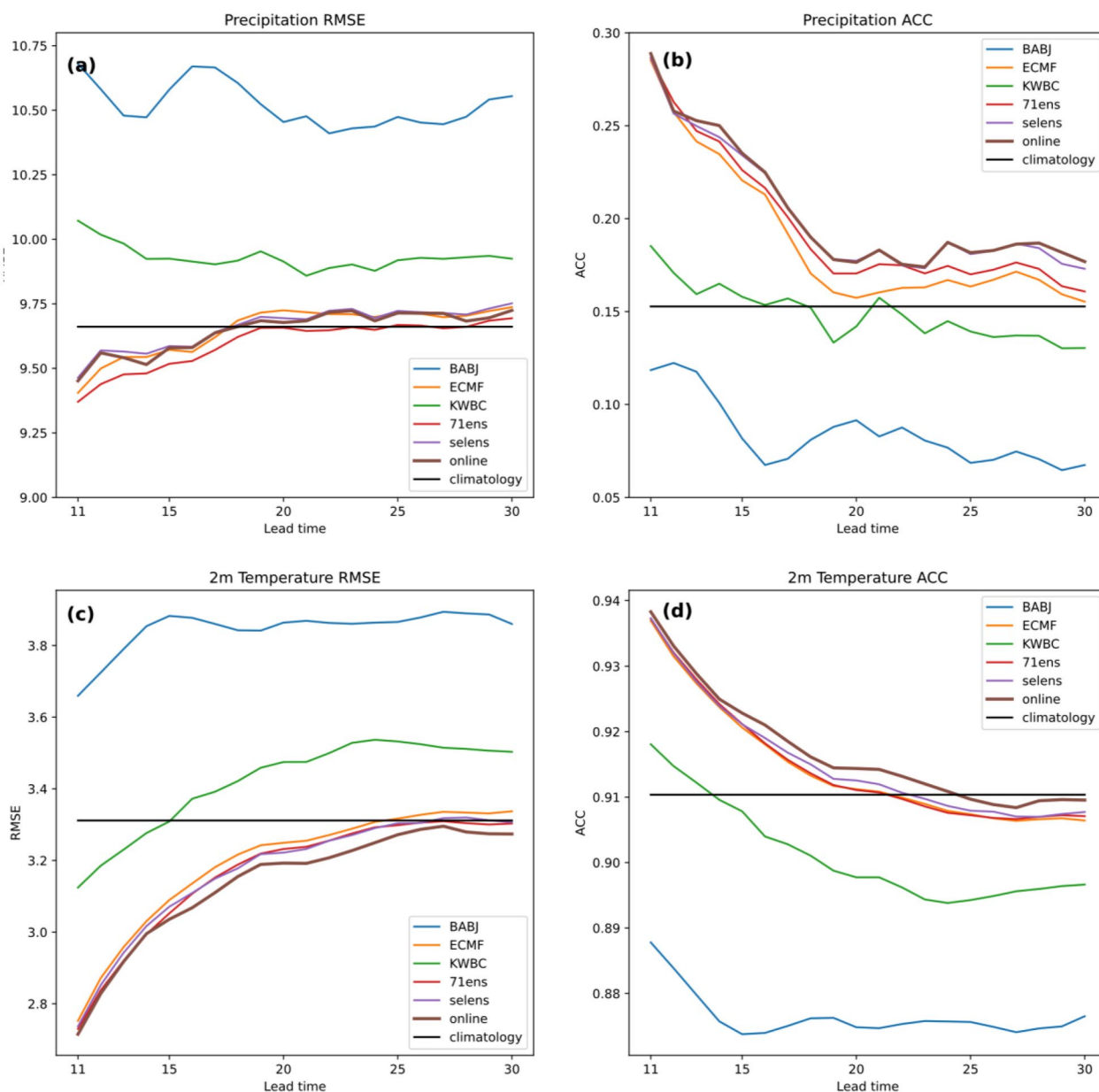


Fig. 5 (a) RMSE (lines; mm) and (b) ACC (lines) for three model ensemble forecasts (blue line for CMA model, green line for NCEP model, yellow line for ECMWF model), equal-weight ensemble of 71 members (red line), equal-weight ensemble of selected members (purple line), online forecasts (brown line) and climatological forecast (black line) for precipitation in the valid set. (c) RMSE (lines; °C) and (d) ACC (lines) for three model ensemble forecasts, equal-weight ensemble of 71 members, online forecasts and climatological forecast for 2 m temperature in the valid set

quite lightweight model, which does not take too much computational resources and time to train compared to neural networks and numerical predictions, this result is satisfactory. Also, when compared against the climatology forecast, we can see that as the lead time increases, the performance for both the ECMWF IFS forecast and the online forecast are getting closer to the climatology prediction, but the precipitation prediction performance

is always better than climatology prediction, which is the same for 2 m temperature prediction except for the lead time of 25–30 days in ACC.

Apart from the comparison based on the hindcasts of the whole valid set for all the lead times, we also selected two typical cases for heavy precipitation and cold wave in 2023 to check the models' ability to predict such extreme events.

Figure 6 shows a case study of forecasting precipitation in the study area for June 24. According the CLDAS record shown in Fig. 6a, there has been heavy rain in most part of Southern China. Xujiahui observatory (31.20°N, 121.43°E) in Shanghai recorded a daily precipitation of 137.9 mm at that day, which was the strongest precipitation in June. Such heavy rainfall in the plum rain season is by far still hard to predict in the extended range forecast and closely connected to economic development and social security. For the precipitation forecast for this specific day, the RMSE for ECMWF forecast, NCEP forecast and online forecast is 26.286, 26.958 and 26.198,

respectively (the advantage over ECMWF forecast is significant on 95% level), with online learning having the best performance. In the comparison shown below, we can see that in real situation in Fig. 6a, the rain belt extended from eastern Guangxi to East China Sea with a southwest–northeast orientation. This characteristic is well demonstrated in the online ensemble forecast in Fig. 6d, while for the other two forecasts, this rain belt is not clearly shown. However, the common problem for ensemble precipitation predictions also appears in this case that although the precipitation area can be clearly shown, the amount is always small. In this case, the

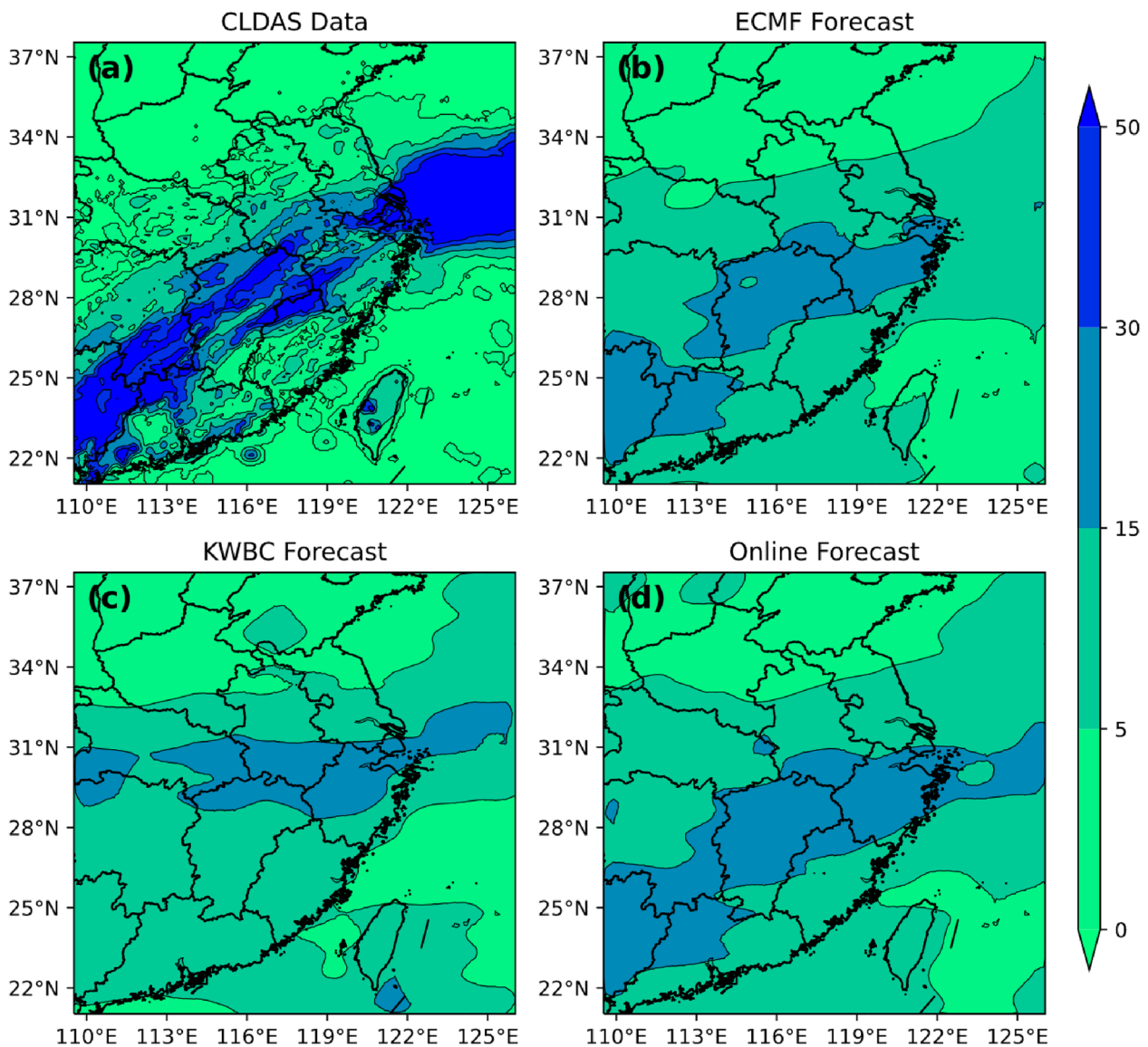


Fig. 6 (a) CLDAS data for the daily precipitation (shading; mm) in 24 June 2023. **b, c** Predictions 14 days ahead by (b) ECMWF IFS, (c) NCEP CFS and (d) online ensemble (shading; mm)

precipitation amount reached 100 mm in some area, but in the predictions, it is hard to surpass 30 mm. Overall, from the perspective of the precipitation area, the online forecast was successful and surpassed the ECMWF and NCEP forecasts.

As the key factors influencing precipitation, the geopotential height and wind on 850 hPa is analyzed for this case. As is shown in Fig. 7, the cyclone over southern Anhui caused the uplift movement of air, together with moisture convergence brought by the convergence of southwesterly to the south of the cyclone, directly caused the rain belt shown in Fig. 6a. To show the physical reliability of our model, we picked 3 model members from the 71 members, which includes KWBC 11 model, ECMF 44

model and BABJ 1 model. The KWBC 11 model takes a relatively big weight of 0.034 while ECMF 44 model takes the smallest of 0.006. The BABJ 1 model is left out in step (2) as a bad model. For the predictions of these models, the KWBC 11 model successfully predicted the cyclone with a small bias in location. The convergence south of the cyclone is also clearly shown in this prediction. The ECMF 44 only predicted a trough over Yangtze River and BABJ 1 model prediction had much bigger bias, with a large cyclone over northern East China Sea. The RMSE for the three predictions of geopotential height at 850 hPa are 31.201, 46.009 and 48.788, respectively, which can prove the physical reliability behind our model. In the establishment of our model, model members with large

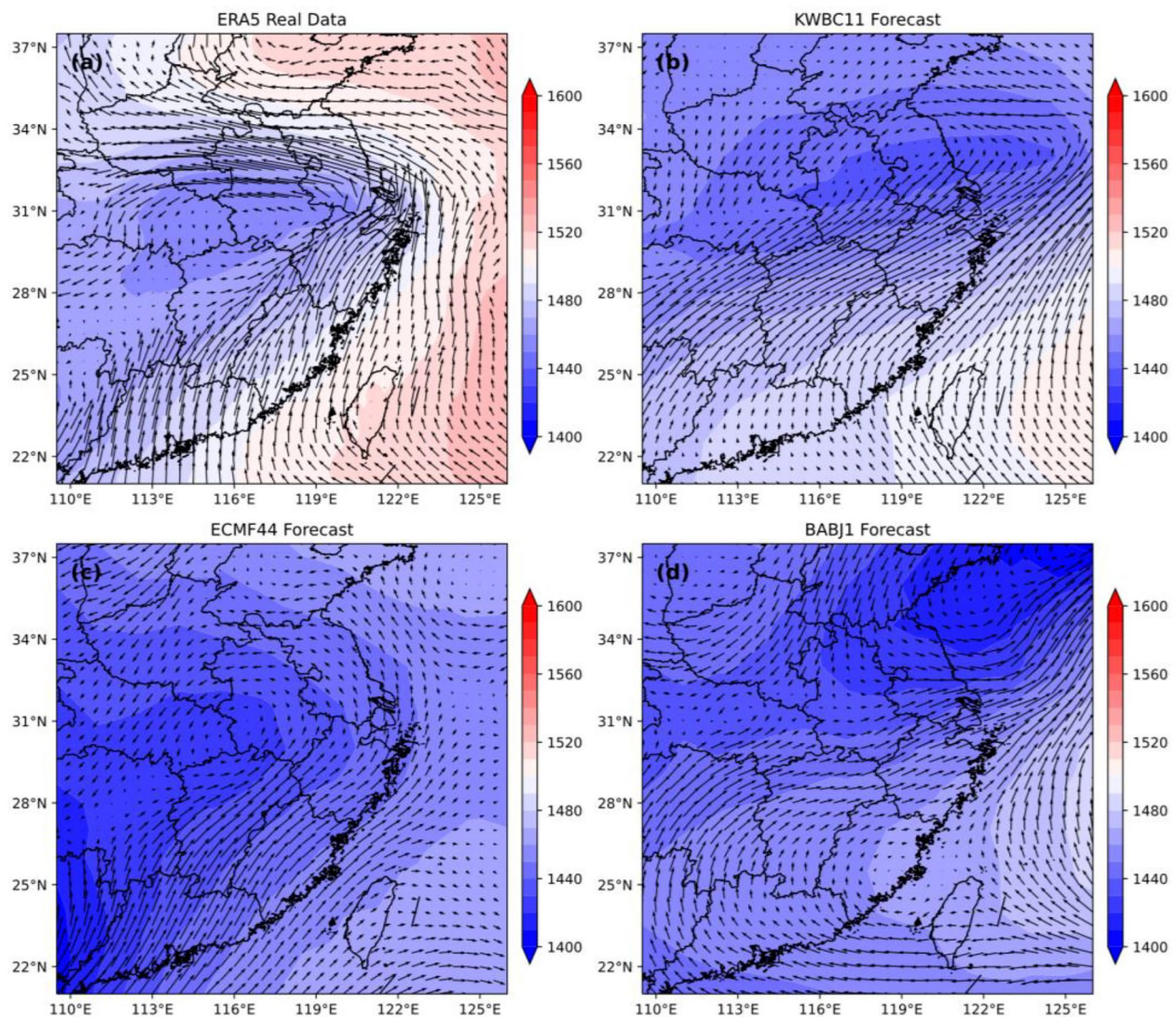


Fig. 7 Geopotential height (shading; gpm) and wind (vector; m s^{-1}) at 850 hPa on June 24, 2023 for (a) ERA5 real data, (b) KWBC 11 model prediction, (c) ECMF 44 model prediction and (d) BABJ 1 model prediction. The model predictions all have the lead time of 14 days

bias are left out in the second step (models like BABJ 1). For the rest of the models, more weight will be given to members with accurate predictions (models like KWBC 11) for the ensemble prediction. The training steps mentioned in Sect. “Data and Methods” ensure that members with accurate predictions can play larger roles in the final ensemble while bad members can be left out and prevented from deteriorating the result.

Another case shown in Fig. 8 is a drastic cold surge happened in 15 January. This cold surge caused the temperature in Southern China to drop by 10.44 °C in 13–15 January during the spring festival travel rush, causing

huge problems to transportation and is thus worth paying attention to. From CLDAS data of 2 m temperature (Fig. 8a), we can see that the area below 5 °C extended to 25°N in 15 January, approaching the coastal area. The RMSE for this day’s forecast is 3.687, 5.593 and 3.357 for ECMWF forecast, NCEP forecast and online forecast, respectively (the advantage over ECMWF forecast is significant on 95% level), which means online learning has the most accurate prediction. Both the ECMWF ensemble forecast and the online forecast predicted the cold area in Eastern China but failed to predict the cold area in the inland area with 30 days in advance, while the

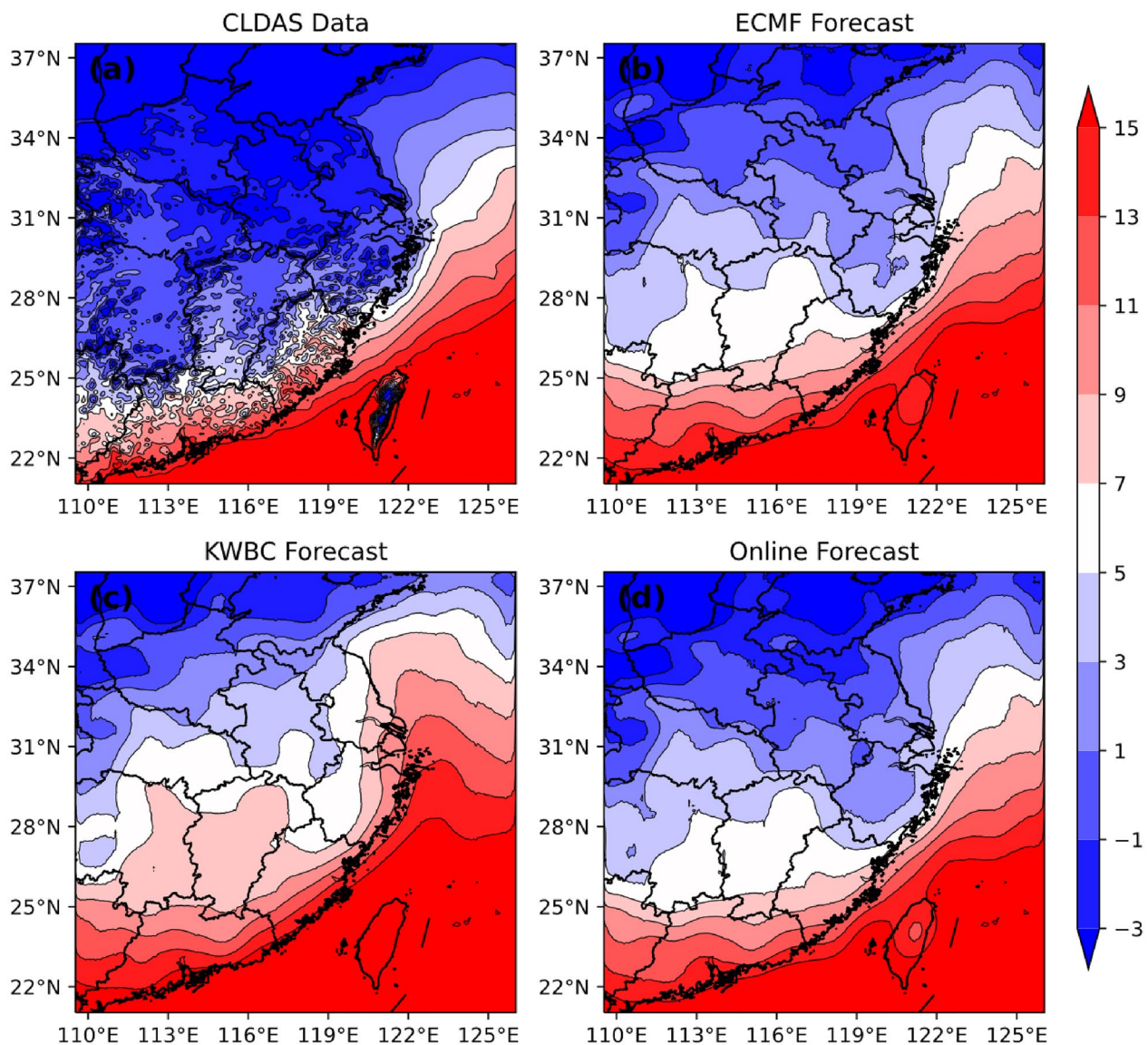


Fig. 8 (a) CLDAS data for the daily average 2 m temperature (shading; °C) in 15 January 2023. **b, c** Predictions 30 days ahead by (b) ECMWF IFS, (c) NCEP CFS and (d) online ensemble (shading; °C)

NCEP ensemble forecast had a larger deviation (Fig. 8c). Compared with the ECMWF forecast (Fig. 8b), the online forecast shown in Fig. 8d had a lower area-average temperature and the cold area are further extended southward, which means it is more in line with the actual situation.

For this case, we also analyzed the circulation background to ensure the physical reliability. As is shown in Fig. 9, South China was under the effect of the short wave troughs which brought cold air southward and lead to cold surges. The depression in Northeast China and the troughs south to it are the critical factors influencing this cold surge. From the model member predictions, we can see that ECMF 31 model, which owns the largest weight of 0.0696, has a good prediction of the circulation background and the depression in the Northeast China is well reflected. For the ECMF 2 model prediction, which has the smallest weight of only 0.0061, has larger bias compared with ECMF 31 model prediction, with the position of the depression being too westward. Finally, the BABJ 3 model, which is already left out in step (2), has the worst prediction, with no troughs but a ridge over South China.

The RMSE for the geopotential height prediction for the three models are 46.788, 110.032, and 113.737, respectively. This case also proves that good models can play a greater role in the final ensemble, which ensure the physical reliability of our model.

Conclusions and discussion

In this passage, we prompted a weighted ensemble forecasting method based on online learning in subseasonal forecast in the SC for precipitation and 2 m temperature. This method is further combined with a modified UNet model for downscaling to produce high resolution predictions. The forecasting results is compared against three model ensemble forecasts and the climatology forecasts through RMSE and ACC as indices and two extreme weather cases. RMSE and ACC shows that this method slightly surpassed ECMWF IFS forecast, which performs the best among the three model ensemble forecasts, and is more accurate than the climatology prediction. In the two extreme weather cases, online learning shows its ability to depict the distribution of atmospheric elements, and its prediction results are more in parallel

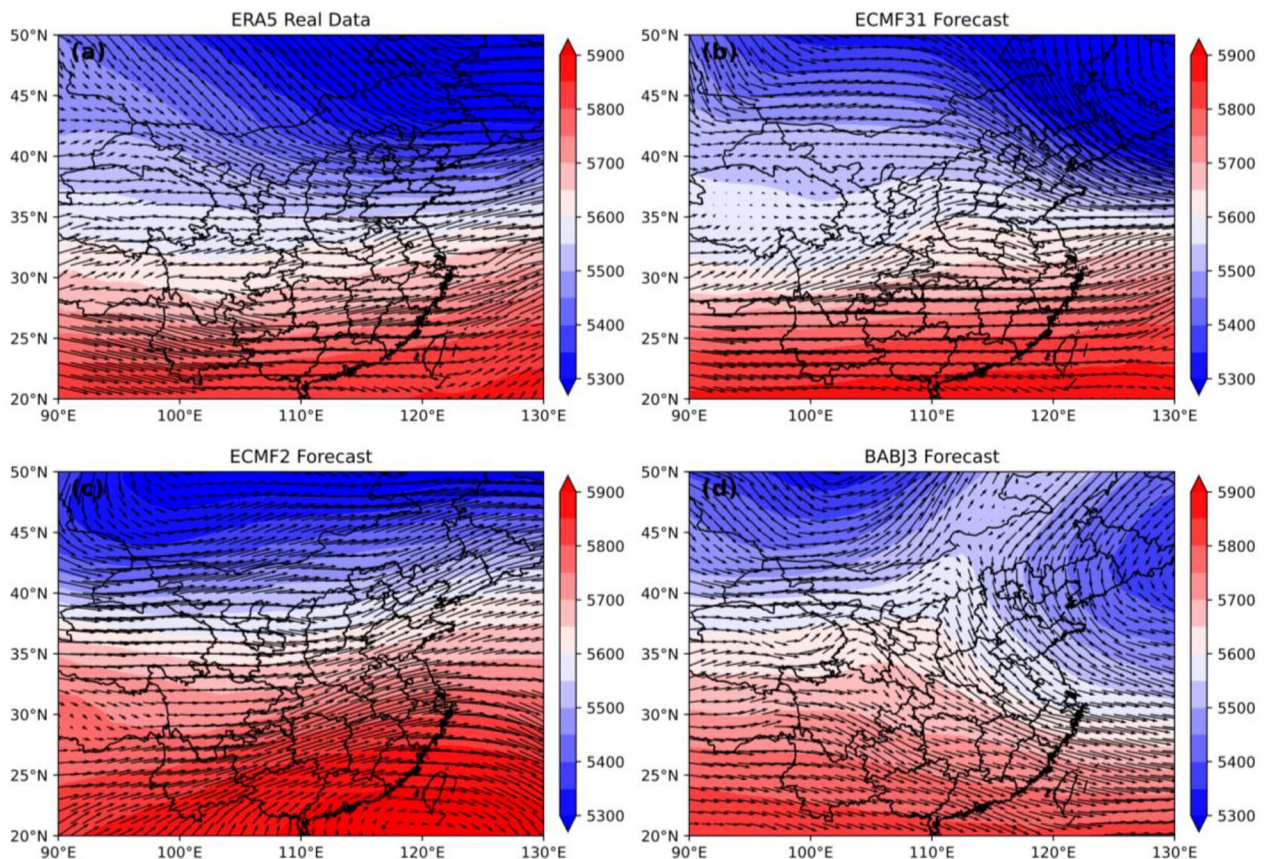


Fig. 9 Geopotential height (shading; gpm) and wind (vector; m s^{-1}) at 500 hPa at January 15, 2023 for (a) ERA5 real data, (b) ECMF 31 model prediction, (c) ECMF 2 model prediction and (d) BABJ 3 model prediction. The model predictions all have the lead time of 30 days

with the CLDAS data. Based on the two cases, we also analyzed the predictions from some of the members and concluded that in our model, the members with less accuracy are filtered out based on their ACC scores. Members with high accurate predictions will be given larger weights so that they can play a greater role in the ensemble prediction and members with relatively low accuracy will be given smaller weights. These are the sources of the improvements our model has gained over other ensemble forecasts.

Compared with neural network predictions, which is quite popular recently, this forecasting method has several advantages. The first is that it does not require GPU resources to train the model, which means all the operations can be done by CPU. The second is that it does not need to take much data to study. In Flaspohlar's passage, the author used the outcomes from a forecast competition with only 200+ samples and still got successful results. For traditional models, we often need to prepare GBs or even TBs of data to train the model, but online learning uses data streams instead of static data sets, gaining this method an advantage. Finally, we have mentioned that for online learning, it can dynamically and real-timely adjust the model according to data feedbacks and thus reflect the changes in input data stream. Unlike neural networks which is based on a static train set and takes much time to update the data set and the model, online learning enables the models to be updated in a higher frequency.

However, this method also has some shortcomings. First, this method makes predictions based on three other model predictions, which means if there is an extreme weather condition that none of the three models manage to forecast, it cannot be predicted by online forecast either. The second problem is already shown in the precipitation prediction part, which is a common problem for all the ensemble forecasts that the amount of precipitation is usually underestimated for heavy rain because there will be ensemble members that predicted the rain and there will also be members that failed to predict. During the ensemble process, the precipitation amount got 'averaged' by the members who failed to predict. Finally, although the RMSE and ACC score we get is slightly better than the ECMWF IFS forecast, they do not pass the 95% level significant test, meaning that this advantage is not significant statistically. These problems remain unsolved in our model and we would possibly be working them out in our future work. Also, the ability of online learning to dynamically and real-timely adjust the model according to data feedbacks is not well demonstrated in this method, improvements around this is also included in our future work.

Acknowledgements

The authors are thankful to the editor and reviewers for their kind feedback.

Author contributions

Conceptualization, FX; methodology, FX; writing—original draft preparation, YS; writing—review & editing, FX, YS and CL; visualization, YS; project administration, FX and CL; funding acquisition, FX. All authors have read and agreed to the published version of the manuscript.

Funding

The study is supported by the Fund of Fujian Key Laboratory of Disaster Weather (2023KFKT06), the National Natural Science Foundation of China (42375067) and Shanghai Science Committee (Grant No. 20dz1200400).

Availability of data and materials

The authors declare that all data are publicly available online including, including the ERA5 Reanalysis Data Set (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels>), the S2S data set (<https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/>) and the CLDAS data set (http://data.cma.cn/data/detail/dataCode/NAFP_CLDAS2.0_NRT.html).

Declarations

Competing interests

The authors have no competing interests to declare.

Received: 27 July 2023 Accepted: 8 January 2024

Published online: 30 January 2024

References

- Baño-Medina J, Manzanar R, Gutiérrez JM (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci Model Dev* 13:2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>
- Bowling M, Burch N, Johanson M, Tammelin O (2015) Headsup limit hold'em poker is solved. *Science* 347(6218):145–149. <https://doi.org/10.1126/science.1259433>
- Flaspohler G, Orabona F, Cohen J, Mouatadid S, Oprescu M, Orenstein P, Mackey LW (2021) Online learning with optimism and delay. *Arxiv*. <https://doi.org/10.4855/arXiv.2106.06885>
- Graepel T, Candela JQ, Borchert T, Herbrich R (2010) Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine. Paper presented at 27th International Conference on Machine Learning ICML, Haifa, Israel, 2010. <https://doi.org/10.5555/3104322.3104326>
- Ham YG, Kim JH, Luo JJ (2019) Deep learning for multi-year ENSO forecasts. *Nature* 573:568–572. <https://doi.org/10.1038/s41586-019-1559-7>
- Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- Hoskins B (2013) The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q J R Meteorol Soc* 139:573–584. <https://doi.org/10.1002/qj.1991>
- Liu X, Yao J, Wu T, Zhang S, Xu F, Zhang L et al (2021) 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*. <https://doi.org/10.1029/2020MS002368>
- Lu C, Kong Y, Guan Z (2020) A mask R-CNN model for reidentifying extratropical cyclones based on quasi-supervised thought. *Sci Rep* 10:15011. <https://doi.org/10.1038/s41598-020-71831-z>
- McMahan HB, Holt G, Sculley D, et al. (2013) Ad Click Prediction: a View from the Trenches. Paper presented at 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Chicago, USA, 2013. <https://doi.org/10.1145/2487575.2488200>
- Orabona F (2019) A modern introduction to online learning. *ArXiv*. <https://doi.org/10.4855/arXiv.1912.13213>
- Roberts CD, Senan R, Molteni F, Boussetta S, Mayer M, Keeley SPE (2018) Climate model configurations of the ECMWF integrated forecasting system

- (ECMWF-IFS cycle 43r1) for HighResMIP. *Geosci Model Dev* 11:3681–3712. <https://doi.org/10.5194/gmd-11-3681-2018>
- Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) *Medical image computing and computer-assisted intervention—MICCAI 2015*. Springer International Publishing, Cham
- Saha S et al (2014) The NCEP climate forecast system version 2. *J Climate* 27:2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Sha Y, Gagne DJ II, West G, Stull R (2020a) Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: daily maximum and minimum 2-m temperature. *J Appl Meteor Climatol* 59:2057–2073. <https://doi.org/10.1175/JAMC-D-20-0057.1>
- Sha Y, Gagne DJ II, West G, Stull R (2020b) Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part II: daily precipitation. *J Appl Meteor Climatol* 59:2075–2092. <https://doi.org/10.1175/JAMC-D-20-0058.1>
- Shi C, Pan Y, Gu J, Xu B, Han S, Zhu Z, Zhang L, Sun S, Jiang Z (2019) A review of multi-source meteorological data fusion products. *Acta Meteorol Sin* 77(4):774–783. <https://doi.org/10.1167/qxxb2019.043>
- Tang S, Qiao S, Feng T, Jia Z, Zang N, Feng G (2020) Predictability of the mid-summer surface air temperature over the Yangtze River valley in the National centers for environmental prediction climate forecast system. *Int J Climatol* 41:811–829. <https://doi.org/10.1002/joc.6670>
- Vitart F et al (2017) The subseasonal to seasonal (S2S) prediction project database. *Bull Amer Meteor Soc* 98:163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>
- Xie J, Hsu P, Hu Y, Ye M, Yu J (2023) Skillful extended-range forecast of rainfall and extreme events in East China based on deep learning. *Wea Forecast* 38:467–486. <https://doi.org/10.1175/WAF-D-22-0132.1>
- Zhang Y, Long M, Chen K et al (2023) Skillful nowcasting of extreme precipitation with NowcastNet. *Nature* 619:526–532. <https://doi.org/10.1038/s41586-023-06184-4>
- Zinkevich M, Johanson M, Bowling MH, Piccione C (2007) Regret minimization in games with incomplete information. In: Platt J, Koller D, Singer Y, Roweis S (eds) *Advances in neural information processing systems*. Curran Associates, Inc., Red Hook
- Zou M, Qiao S, Yang Y et al (2022) Predictability of the two temperature modes of the East Asian winter monsoon in the NCEP-CFSv2 and MRI-CPSv2 models. *Clim Dyn* 59:3211–3225. <https://doi.org/10.1007/s00382-022-06254-9>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.