



Improving the forecast accuracy of ECMWF 2-m air temperature using a historical dataset

Zhaolu Hou ^{a,b}, Jianping Li ^{a,b,*}, Lei Wang ^c, Yazhou Zhang ^{a,b,1}, Ting Liu ^{d,e}

^a Frontiers Science Center for Deep Ocean Multispheres and Earth System (FDOMES)/Key Laboratory of Physical Oceanography/Academy of the Future Ocean/College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China

^b Laboratory for Ocean Dynamics and Climate, Pilot Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

^c Beijing Presky Technology Co., Ltd., Beijing 100089, China

^d State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

^e Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China



ARTICLE INFO

Keywords:

Model forecast correction
Analog
Historical data
2-m air temperature
Cold wave
Qinghai Tibet Plateau

ABSTRACT

The 2-m air temperature (T2m) is an important meteorological variable and has been the focus of meteorological forecasting. Although the numerical weather model is an important means of forecasting, it typically presents forecasting errors that cannot be eliminated by improving the ability of the numerical model to reproduce the processes. Thus, a statistical correction of the forecast results is required. In this study, we applied the local dynamical analog (LDA) method to correct the operational T2m forecast product obtained from the European Centre for Medium-Range Weather Forecasts with the lead time of 24–240 h. To our knowledge, for the first time, we used spatially adjacent grids from high-resolution grid data as potential analog pools to compensate for the short duration of historical data. The T2m of weather forecasts in East Asia for December 2018 was improved by LDA correction with a small sample condition. Compared with ERA5 and station observation data, the results show that the root mean square error can be reduced by 2%–4% and the correlation coefficient can be increased by 1%–5% for different lead times, with the most distinct improvement effect for the medium-term forecast time. The Qinghai Tibet Plateau, Mongolia Plateau, and other areas, where the raw prediction error is relatively high, presented better performance than other regions. For a cold-wave process, we also demonstrate that the corrected results based on analogs present better forecasting skill performance than raw forecast results. The analog correction with the LDA method, which combines statistical and model dynamical techniques, is proposed to be integrated with other advanced operational models. The forecast skill of T2m was improved by a historical dataset, which may contribute to energy management and the construction industry.

1. Introduction

The 2-m air temperature (T2m) is an important meteorological variable that affects daily human life. Determination of T2m is necessary to estimate the performance of solar heating and photovoltaic solar energy in energy applications as well as for the construction industry, including the design of building heating and cooling systems (Clarke, 2007; Huld et al., 2010; Koehl et al., 2011). Thus, T2m is one of the most important output parameters from the numerical weather prediction (NWP) model and is forecasted daily. Its accuracy has received significant attention from users of the forecast results. Moreover, T2m is a part

of the standard verification process of the NWP model performance and is an important competition parameter with respect to other weather forecast providers. Although the NWP model is undergoing constant development and improvement (Eyre et al., 2020; Lynch, 2006; McTaggart-Cowan et al., 2019; Shen et al., 2020; Thomas et al., 2020), T2m modeling typically presents forecast errors under certain meteorological and terrain conditions for the verification of model forecasts, for example, cold-wave weather and mountain terrain. Forecast errors result from several factors; for example, under complex terrain, parameterization of land surface processes is difficult using a relatively few initial observation data with poor representativeness (Bauer et al.,

* Corresponding author at: Frontiers Science Center for Deep Ocean Multispheres and Earth System (FDOMES), Key Laboratory of Physical Oceanography, Academy of the Future Ocean, College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China.

E-mail address: ljp@ouc.edu.cn (J. Li).

¹ Co-first author of this work.

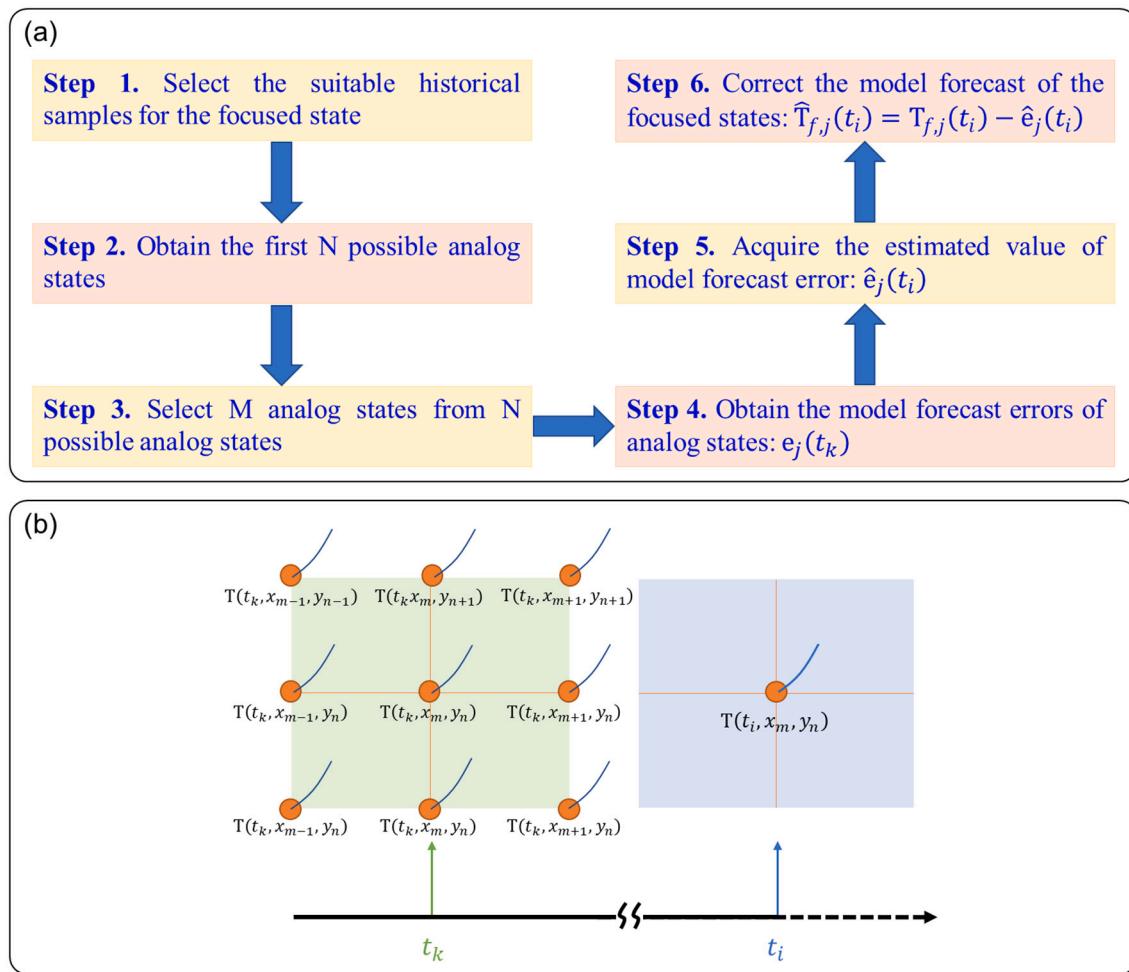


Fig. 1. (a) Flow chart for the LDA-correction method. (b) Schematic of model forecast correction to locate analogs with adjacent grids. $T(t_i)$ is the focused reanalysis/observation state at the time t_i . $T_{f,j}(t_i)$ is the model forecast value at the lead time of j from $T(t_i)$. $e_j(t_k)$ is the model forecast error of analog states $T(t_k)$ forecasting from the initial time t_k to time t_{k+j} . $\hat{e}_j(t_i)$, as the estimation of $e_j(t_i)$, is obtained by fitting $e_j(t_k)$. In the focused state $T(t_i, x_m, y_n)$ from Fig. 1b, x denotes the latitude grid point and y is the longitude grid point. $T(t_k, x_m, y_n)$ and $T(t_k, x_{m\pm1}, y_{n\pm1})$ indicate the possible analog and its adjacent grid points of the focused state at the time t_k .

2015; Kam et al., 2021). Post-processing methods are excellent statistical methods for improving the performance of NWP models, as they are relatively simple to apply and computationally inexpensive (Lorenc, 1986; Sweeney et al., 2013).

Post-processing methods for reducing model forecast errors can be classified as state-independent and state-dependent methods (Danforth and Kalnay, 2008). Generally, statistical methods used to correct model forecast errors are termed as state-independent corrections, such as model output statistics (MOS) (Carter et al., 1989; Glahn and Lowry, 1972). MOS is a commonly used method that mainly uses observation data to match the output of NWP and then obtains a prediction equation based on linear regression (Glahn and Lowry, 1972; Marzban et al., 2006). MOS as state-independent corrections are linear methods that can only reduce the systematic component of model forecast errors (DelSole and Hou, 1999; DelSole et al., 2008), which have certain limitations owing to the potentially nonlinear relationship between predictands and predictors. Simultaneously, MOS correction approaches often focus on correcting forecast data at individual weather observation stations. Refined grid forecasting is becoming increasingly important in weather prediction models; however, state-dependent components constitute most of the total model forecast errors (Dalcher and Kalnay, 1987). State-dependent corrections are, therefore, required to reduce the state-dependent components and further improve the forecast accuracy. Thus, the development of grid point-based forecast-bias correction techniques that reduce state-dependent forecast error is

important and presents an urgent challenge (Vannitsem et al., 2020).

In addition, machine learning techniques have received considerable attention recently across several fields, including atmospheric science (Rasp and Lerch, 2018; Boukabara et al., 2019; Chen et al., 2019). Ji et al. (2019) reported that artificial intelligence (AI)-based ensemble models exhibited improved skill in forecasting surface air temperature. Likewise, Peng et al. (2020) applied two machine learning methods to the 2-m maximum air temperature forecast with lead times of 1–35 days over East Asia and showed that all the post-processing methods can efficiently reduce prediction biases and uncertainties, especially in the lead weeks 1–2. Reasonable machine learning techniques as post-processing methods were found to contribute to obtaining better forecasts. However, these studies paid little attention to T2m forecast correction in the lead time of 24 to 240 h, and machine learning or AI-based techniques usually correspond to computational complexity and require a large amount of data. Furthermore, the impact of time dimension, including the dynamic evolution information, was often ignored in machine learning correction.

If the forecast state is regarded as a small disturbance superimposed on an analogous historical state, the similarities of the model forecast errors between the trajectories of analogs can be exploited, and the model forecast errors can be corrected efficiently using an analog-dynamical method. A key aspect of such a method is the identification of analogs. Hou et al. (2020, 2021) proposed a new correction method for model forecast errors based on the local dynamical analog

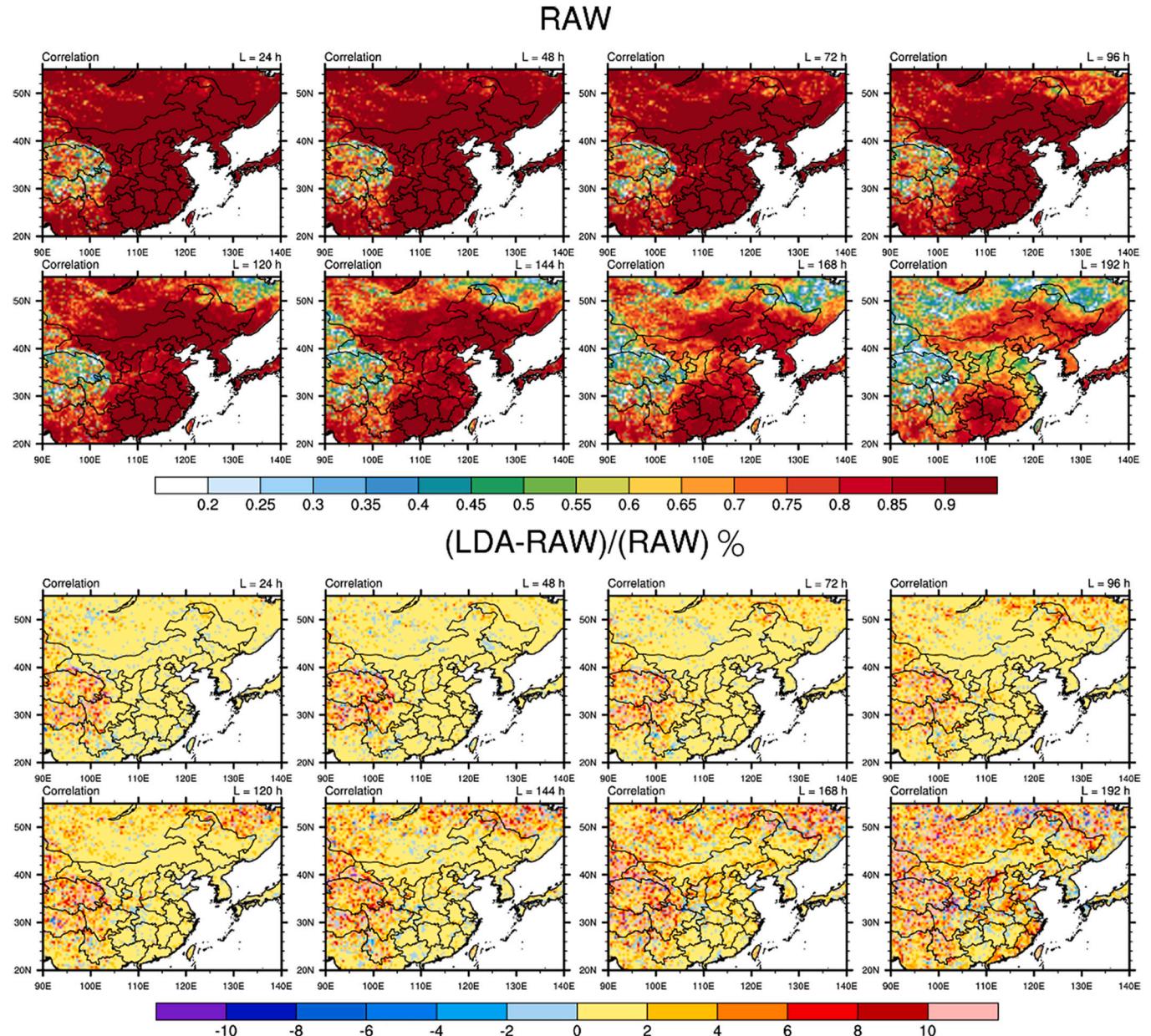


Fig. 2. Temporal correlation coefficients between ERA5 and ECMWF daily forecast 2-m air temperature in December 2018. The first two rows represent the correlation skill of ECMWF raw forecast for the lead time of 24, 48, 72, 96, 120, 144, 168, and 192 h. The last two rows indicate the improvement percentage of the correlation skill from the LDA-correction method for different lead times. The improvement percentage is calculated by $\frac{\text{corr}(\text{corrected}) - \text{corr}(\text{raw})}{\text{corr}(\text{raw})} \times 100\%$, where $\text{corr}(\text{corrected})$ is the correlation skill between the corrected forecast result and reanalysis/observation and $\text{corr}(\text{raw})$ is the correlation skill between the raw forecast result and observation in every grid. The positive improvement percentage represents the increase in correlation skill.

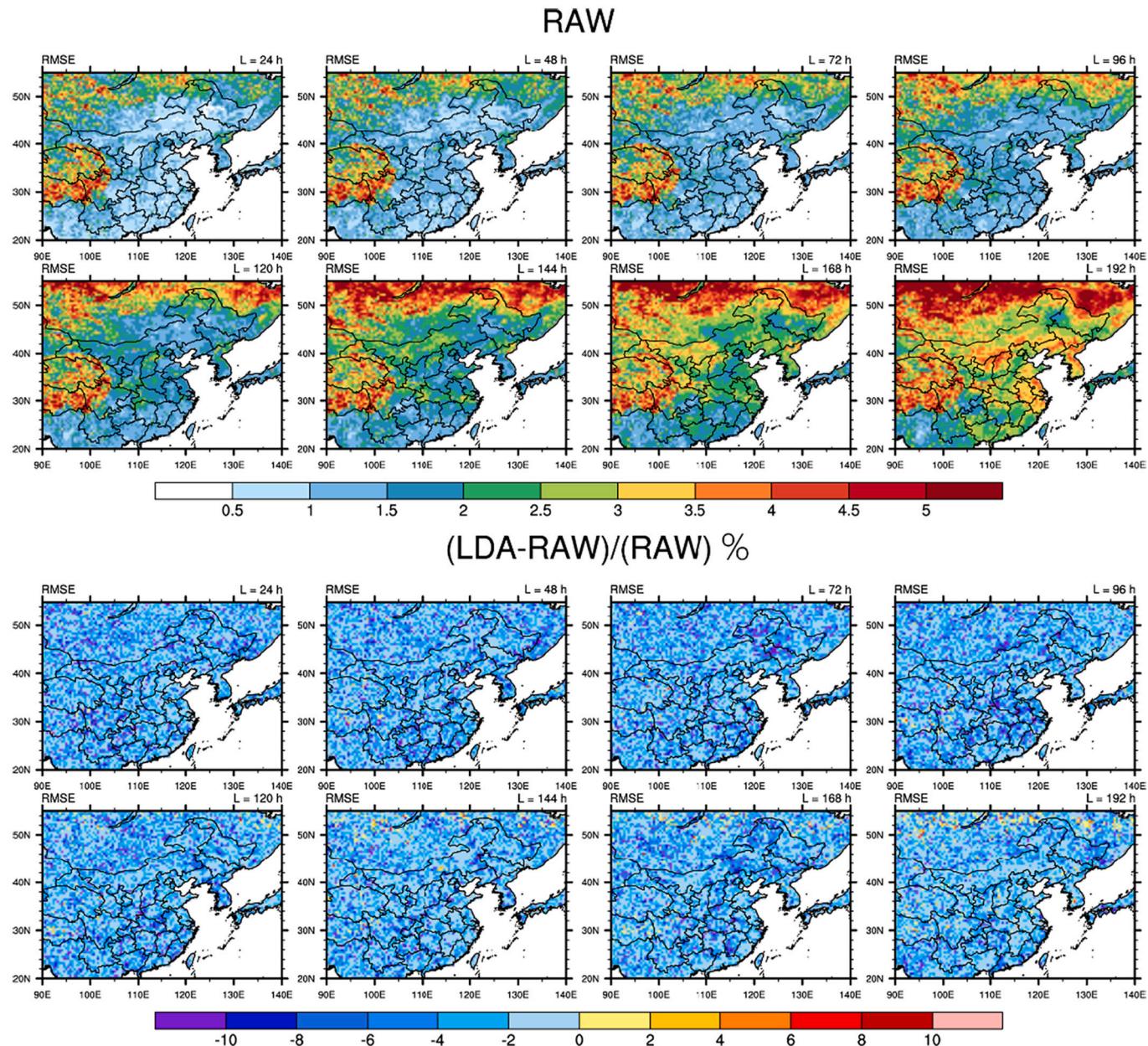


Fig. 3. Root mean square error (RMSE) between ERA5 and ECMWF daily forecast 2 m air temperature in December 2018. The first two rows represent RMSE skill of the ECMWF raw forecast for the lead time of 24, 48, 72, 96, 120, 144, 168, and 192 h. The last two rows indicate the improvement percentage of the RMSE skill from the LDA-correction method for different lead times. The improvement percentage is calculated by $\frac{\text{RMSE}(\text{corrected}) - \text{RMSE}(\text{raw})}{\text{RMSE}(\text{raw})} \times 100\%$, where RMSE(corrected) is the RMSE skill between the corrected forecast result and reanalysis/observation and RMSE(raw) is the RMSE skill between the raw forecast result and observation in every grid. The negative improvement percentage represents the decrease in RMSE skill.

(LDA) method with a historical analog dataset. The LDA method is an analog-locating method that ensures the similarity between the dynamic evolution of both states, rather than only considering their initial states (Ding et al., 2016; Ding and Li, 2007; Li and Ding, 2011, 2015; Li and Wang, 2008). Based on the historical forecast dataset, this method reflects the nonlinear relationship between the forecast error and the initial state. As a state-dependent and nonlinear correction method, the LDA-correction method is a grid point-based forecast-bias-correction technique that has been successfully applied to seasonal prediction products from different models, such as an intermediate coupled model of the El Niño–Southern Oscillation and Climate forecast system Version 2 (Hou et al., 2020, 2021). These results have substantiated that the LDA-correction method is a promising method for application in existing numerical models and improving forecast skill. In contrast to seasonal

prediction, the weather forecast is closely related to human life and health; weather forecast model versions are updated frequently, which leads to the short length of historical data. Therefore, the application of this method to medium-range forecasting of T2m with high-resolution and short-time datasets still needs to be solved. Additionally, refined grid forecasting is becoming increasingly important in weather prediction models; however, the existing correction approaches pay most attention to correcting forecast data at individual weather observation stations (Cho et al., 2020; Marzban, 2003; Vislocky and Young, 1989). Therefore, the development of grid point-based forecast-bias correction techniques is an urgent challenge (Vannitsem et al., 2021); it is essential to study the application of the LDA-correction method to weather operational forecasts and evaluate its performance.

Thus, we applied the LDA-correction method to the analog

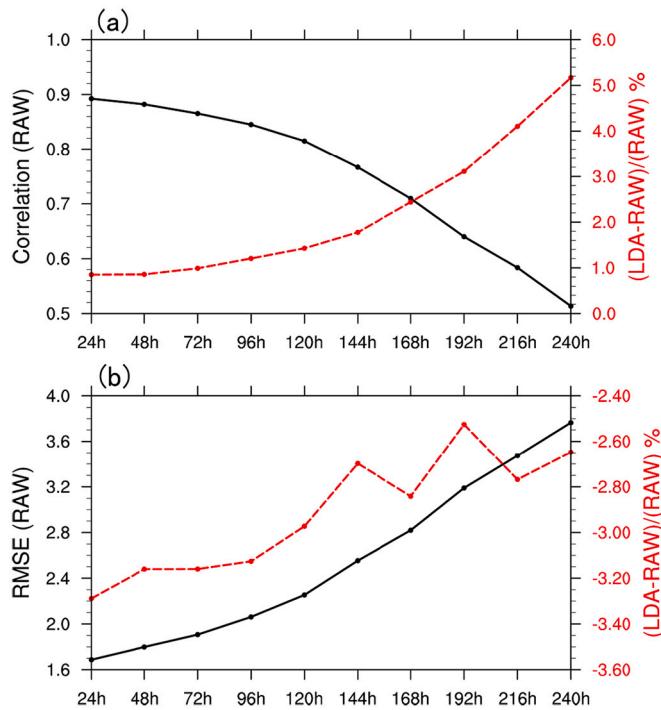


Fig. 4. Spatial average results of temporal correlation (a) and RMSE skill (b) over the study region (20° – 55° N, 90° – 140° E) for different lead times from 24 h to 240 h based on Figs. 1 and 2. The black lines (left vertical coordinates) show the raw forecast skill of ECMWF, and the red curves (right vertical coordinates) indicate the improvement percentage from the LDA-correction method relative to raw forecast skill for different lead times. The spatial average value is based on the cosine of the latitudes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

correction of T2m forecast results from state-of-the-art operational weather models, focusing on the correction performance from complex terrains and short historical datasets. However, applying the LDA-correction method to weather forecast results of small samples with high resolution presents a problem, which we solved by updating the LDA-correction method in this study. When locating analog states of the focused state, we used the forecast results of the adjacent grid points as historical data, an approach that differs from that of previous research. Moreover, spatially adjacent grid states have certain common characteristics in forecast error, which compensates for the deficiency of historical data in the weather forecast and contributes to reducing forecast errors of the interest states by the LDA-correction method. The performance of the LDA-correction method was evaluated for the small sample condition. The method and data used are introduced in Section 2 and the results of the operational weather models are described in Section 3. The conclusion and discussion are presented in Section 4.

2. Method and data

2.1. Data

We applied the LDA-correction method to the T2m operational weather forecasts and selected East Asia (20° – 55° N, 90° – 140° E) as the study area, considering its variable underlying surface types and large population. For the operational weather forecast, we selected the European Centre for Medium-Range Weather Forecasts (ECMWF) operational atmospheric model high-resolution 10-day forecast results for December 2018. The horizontal resolution of these data was $0.125^{\circ} \times 0.125^{\circ}$ (latitude \times longitude). The daily T2m from this forecast product was corrected at 00:00 UTC for each day. The forecast results from December 9 and 11 were eliminated due to missing forecasts. The

verification data were the hourly T2m from the ERA5 reanalysis data. ERA5 is the fifth generation of the ECMWF atmospheric reanalysis of global climate and combines vast amounts of historical observations into global estimates using advanced modeling and data assimilation systems (Hadden et al., 2018). The hourly T2m from the ERA5 reanalysis data was $0.1^{\circ} \times 0.1^{\circ}$; thus, it was interpolated to $0.125^{\circ} \times 0.125^{\circ}$.

Station observation is independent from ECMWF forecast product for validation. Thus, we used station data from the Integrated Surface Dataset (NOAA National Centers for Environmental Information, 2001). We filtered out the observational stations in the focused region (20° – 55° N, 90° – 140° E) with existing valid values in December 2018, resulting in 410 stations passing the quality control. Subsequently, we used the most used bilinear interpolation method to interpolate grid forecast data to the latitude and longitude of the chosen observation stations and corrected these forecast results using the LDA-correction method.

Prior to being evaluated and corrected by using the LDA-correction method, the model forecast result was processed using the following operation: setting the missing value. Grids with half missing forecast data at those time points were set as missing values and were not considered in subsequent corrections.

2.2. Method

The LDA-correction method is an analog-based correction method, which takes advantage of the model forecast error of analogous states in historical data (that are retrieved in advance) to correct the forecast results of the state of interest whose subsequent actual conditions have not yet been observed. The LDA method is an analog-locating method that not only considers similarities in their initial spatial structures but also includes dynamical evolution information (Gao et al., 2006; Li and Ding, 2011; Li and Ding, 2015). Therefore, this method can best invert the error growth structure. The advantages and operation steps of the LDA-correction method are described by Hou et al. (2020, 2021).

In this study, the LDA-correction method was applied to correct the ECMWF T2m forecast product. The model forecast T2m is a field variable. Here, we located analog states using T2m time series in every grid as $T(t_i)$. The T2m time series at each grid was considered. The main process and steps of the LDA-correction method were (Fig. 1a):

Step 1. Select the suitable historical samples for the focused state:

For every focused state, the suitable historical samples were limited to historical states from adjacent grid data with the same hour in the day. Since spatially adjacent grid states have certain common characteristics in forecast error, we used the forecast results of adjacent grid points as historical data when locating analog states of the focused state (shown in Fig. 1b), as opposed to previous research.

Step 2. Obtain the first N possible analog states:

For every suitable historical state $T(t_k)$, we calculated the total distance d_t between state point $T(t_i)$ and $T(t_k)$ based on the reanalysis/observation dataset. d_t is defined by adding the initial distance $d_i = \|T(t_i) - T(t_k)\|$ and the evolutional distance d_e . d_e is based on the local distance before the initial states with unequal weight coefficients as (Hou et al., 2021):

$$d_e = \sqrt{\frac{1}{L} \sum_{j=1}^L |T(t_{i-j}) - T(t_{k-j})|^2 \left(\frac{RMSE(1)}{RMSE(j)} \right)^2} \quad (1)$$

where $RMSE(j)$ is the root mean square error (RMSE) of persistent forecast at the lead time of j month compared with observation. We sorted these potential states from small to large based on the distance d_t of the LDA-correction method and used N analogs with the smallest distance as the potential analogs.

Step 3. Select M analog states from N possible analog states:

The forecast results of each analog states are known, and the raw forecast results of the focused state were also obtained before correcting

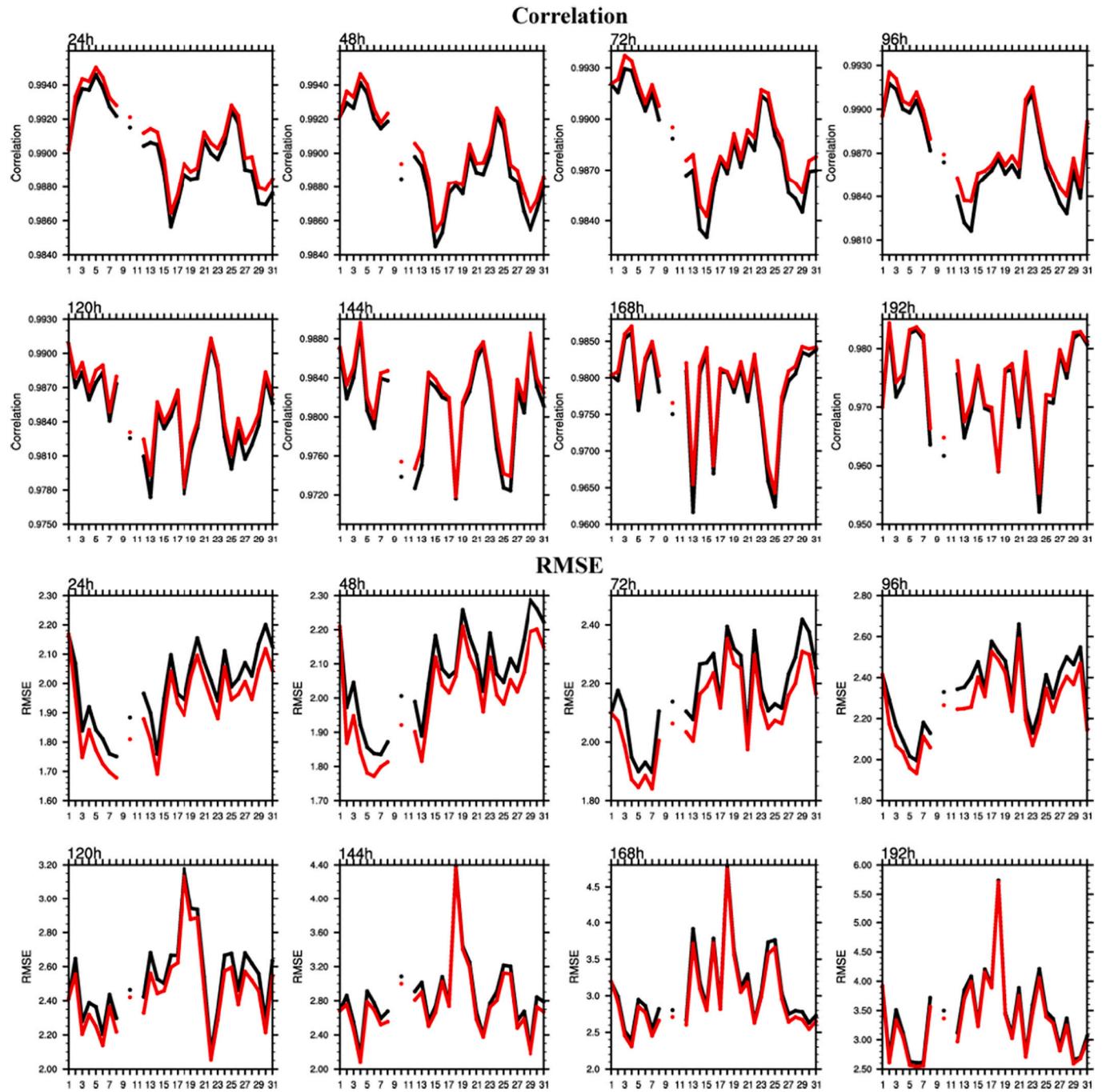


Fig. 5. Spatial correlation coefficient (the first two rows) and RMSE skill (the last two rows) from 24 h to 192 h of lead time for every day of December 2018. The black lines are the forecasting skill from the raw forecast, and the red lines present the results from the corrected forecast results by using the LDA-correction method. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

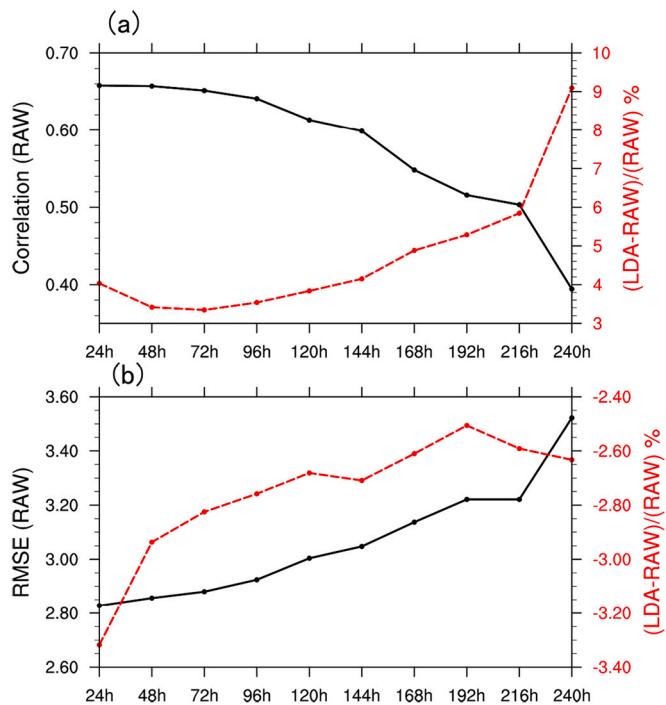


Fig. 6. Spatial average results of temporal correlation (a) and RMSE skill (b) for the Qinghai Tibet Plateau (26° – 39° N, 90° – 104° E) for different lead times from 24 h to 240 h based on Figs. 1 and 2. The black lines (left vertical coordinates) show the raw forecast skill of ECMWF, and the red curves (right vertical coordinates) indicate the improvement percentage from using the LDA correction method relative to raw forecast skill for different lead times. The spatial average value is based on the cosine of the latitudes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the forecast bias. We calculated the correlation coefficient between the analogs' forecast results and the focused state's forecast results during the forecast period of F_h . Based on the correlation coefficient, we chose M analog states possessing the highest values from N potential analogs as the final analog states and subsequently derived their average forecast error.

Step 4. Obtain the model forecast errors of analog states:

Following derivation of the analog $T(t_k)$ of the focused states $T(t_i)$, model forecast error forecasting from the initial time t_k to time t_{k+j} can be described as follows:

$$e_j(t_k) = T_{f,j}(t_k) - T(t_{k+j}) \quad (2)$$

where $T_{f,j}(t_k)$ is the model forecast value at the lead time of j from $T(t_k)$, f represents the forecast result.

Step 5. Acquire the estimated value of model forecast error:

When M final analog states are considered, the mean analog forecast error $\bar{e}_j(t_k)$ is $\frac{\sum_{m=1}^M e_j(t_{km})}{M}$ and the estimation of $e_j(t_i)$ can be described as follows:

$$\hat{e}_j(t_i) = \alpha_j \frac{\sum_{m=1}^M e_j(t_{km})}{M} + \beta_j \quad (3)$$

where α_j and β_j are parameters in the lead time j which can be determined by using the linear regression based on the training data.

Step 6. Correcting the model forecast of the focused states:

$e_j(t_i)$ can be partially eliminated by $\hat{e}_j(t_i)$, and the model forecast T_f , $T(t_i)$ can be corrected as:

$$\hat{T}_{f,j}(t_i) = T_{f,j}(t_i) - \hat{e}_j(t_i) \quad (4)$$

Some key parameters were set for the LDA-correction method. First, the window span L in the LDA method was related to the autocorrelation coefficient of the data series (Li and Ding, 2011, 2013) and set as $L = 24$ h, which includes eight points because of 3-h interval. Considering the high spatial resolution of the ECMWF forecast product, we considered the nine adjacent grid data series near every focused state as potentially similar states (Fig. 1b). During the operation, we calculated the LDA distance for every potential analog. Then, we arranged these potential states from small to large. We used $N = 3$ analogs with the smallest distance as the chosen analogs. Based on the forecast direction consistency principle, we selected $M = 2$ analogs and obtained their average forecast errors. α_j and β_j are parameters in the lead time j , which can be determined by using linear regression based on the training data. Owing to the limited amount of data, the training period was chosen by using sliding selection, and 48 h were checked off from the time of focused states. For example, when correcting the forecast results initialized from December 12, 2018, the forecast data from ECMWF and observational data of the periods from December 1, 2018, to December 10, 2018, and from December 14, 2018, to December 31, 2018, was regarded as the training dataset. Based on the training dataset, we derived the real forecast errors and their averaged analog errors for every state. The line regression parameter was evaluated by both forecast errors. Based on the known forecast error of analogs, the forecast results initialized from the focused state were corrected using the LDA-correction method.

3. Results

We corrected the ECMWF T2m forecast product for December 2018 using the LDA-correction method and focused on daily forecast results from 00:00 UTC. Since medium-range weather forecasts are important for both society and economy, the forecast lead times covered 24–240 h with 24-h intervals. Thus, the total number of initial forecast times was 29. The ERA5 reanalysis data for T2m were used as the observations. The configuration was designed to confirm the applicability of the LDA-correction method in the case of small samples. The LDA-correction method selected three potential analogs for a single focused state and used the first two to obtain the analog model forecast error according to the correlation of the forecast trend. The window length used for the LDA method was 24 h. In winter, East Asia often suffers from cold-wave events; therefore, the forecast quality of air temperature is of great concern (Kim et al., 2012; Tian et al., 2018; Zhu and Li, 2017).

To exhibit the forecast skill of T2m from ECMWF for different lead times, Fig. 2 shows the temporal correlation coefficients between the observations and model forecasts of T2m during December 1–31, 2018, at lead times of 24, 48, 72, 96, 120, 144, 168, and 192 h. The first two rows showcase the forecasting skill based on the raw forecasts. The correlation forecast skill of T2m was high and exceeded 0.75 in most areas in the early lead time. With an increase in lead time, the forecast skill decreased. Compared to eastern and southern China, the Qinghai Tibet Plateau and Mongolian Plateau had low correlation, representing a deficiency in the ECMWF operational weather raw forecast product, which may respond to complex terrain and few observation data. With the application of the LDA-correction method, the forecast performance of T2m was improved. The last two rows display the improvement percentages relative to the raw forecast skill. Positive values represent the improved temporal correlation coefficients. Overall, the proportion of improvement of spatial grids reached more than 80% at different lead times. The regions where the raw forecast presented low correlation coefficient correspond to areas showing great improvement by using the LDA-correction method, such as the Tibetan Plateau, Qilian Mountains, and Mongolian Plateau. In these regions, the improvement reached 10% or even 20% of the raw forecast correlation coefficient. This percentage improvement held for all lead times from 24 h to 240 h. Fig. 4a shows the spatial average results of the temporal correlation coefficients for the raw forecast and the improvement percentage based on Fig. 2. The mean improvement percentage was close to 1% for the lead time of 24–96 h

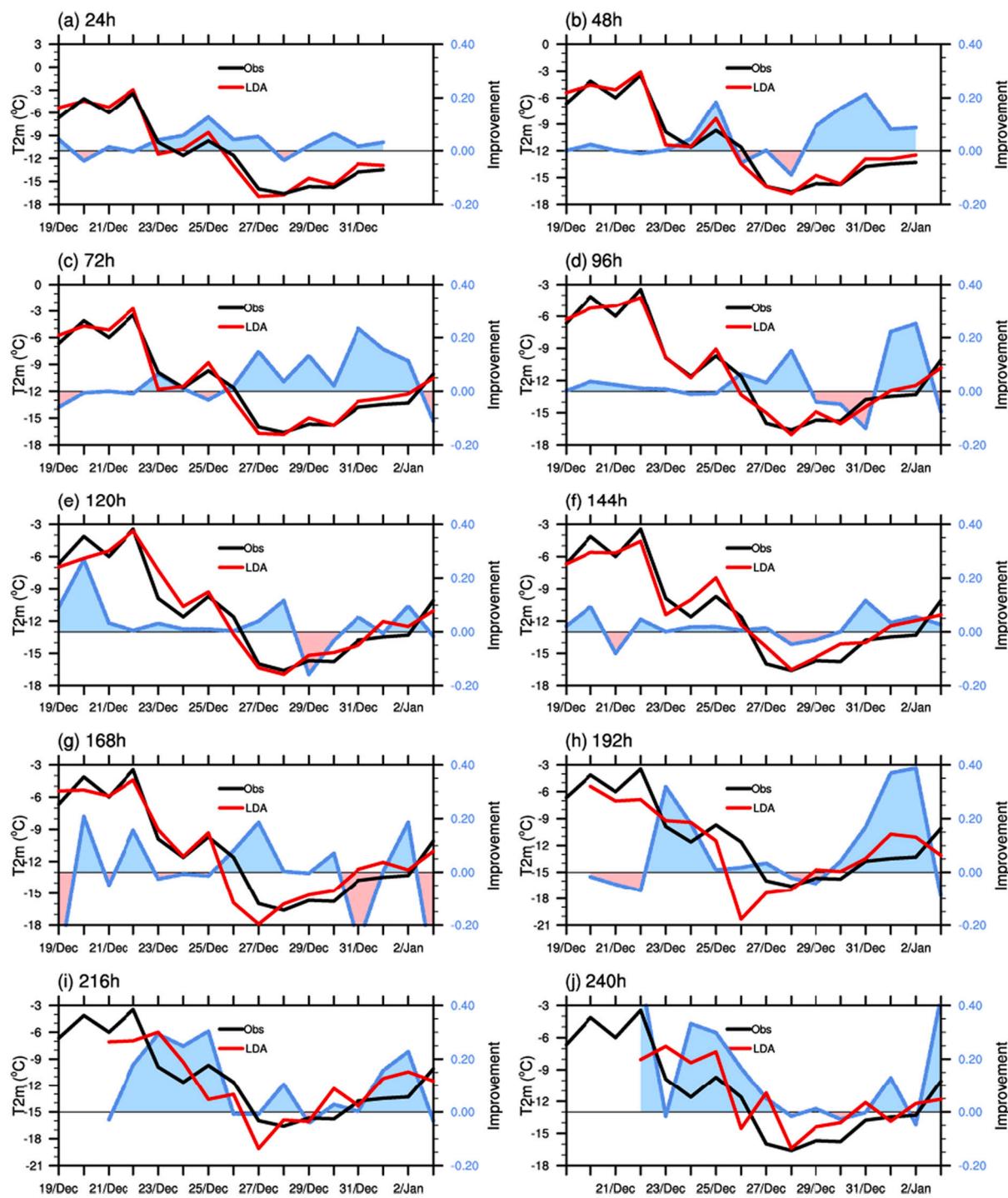


Fig. 7. Corrected forecast results by using the LDA-correction method (red line) and reanalysis/observational (black line) data for 2-m air temperature over Beijing ($39.5^{\circ}\text{--}41.0^{\circ}\text{N}$, $115.5^{\circ}\text{--}117.5^{\circ}\text{E}$) in the lead time of 24 h to 240 h. Blue shading represents reduced error of the LDA-corrected forecasts compared with the raw forecast. Light-pink-shaded values is the time when the LDA-correction method does not improve the forecast. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

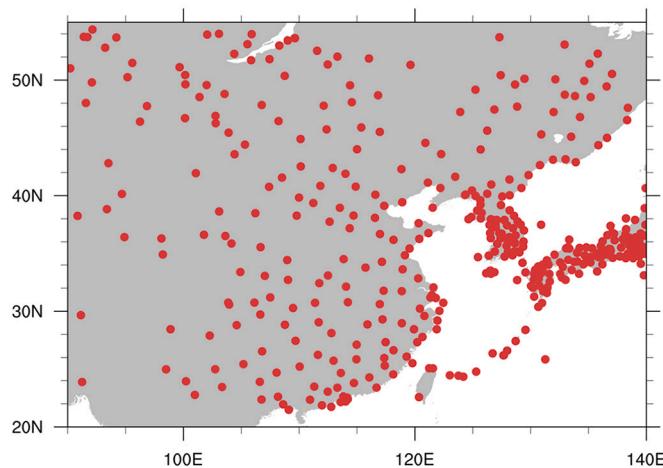


Fig. 8. Observation stations over the regions (20° – 55° N, 90° – 140° E) used to obtain the validation data from the Integrated Surface Dataset (NOAA National Centers for Environmental Information). The total number of used stations is 410.

and increased to 5% with an increase in lead time to 240 h, which showcases the effect of improvement due to the LDA-correction method.

The advantage of the LDA-correction forecasts over the ECMWF raw forecast was indicated by the RMSE skill (Fig. 3). For the raw forecast results, the RMSE values were small over eastern China and large in the Qinghai Tibet Plateau and Mongolian Plateau. Along with the increase in lead time, RMSE values increased, representing a decrease in forecast accuracy. The improvement percentage relative to the RMSE from the raw forecasts is shown in the last two rows of Fig. 3. The negative values indicate that the LDA-correction method reduced the RMSE of the ECMWF forecast across almost the entire focused region. The ratio of improved grid numbers in the focused region was greater than 92% for the lead time of 24–240 h. In the Tibetan Plateau region and Mongolian Plateau, LDA correction reduced the RMSE by 0.15°C . The improvement in RMSE reached over 4% of the RMSE value of the raw model forecast in the region of interest and nearly 10% for certain points (not shown). Fig. 4b shows the spatial average results of the RMSE based on Fig. 3. In the focused region, the mean reduction percentage of RMSE reached more than 2.5%, reaching even 3.5% at a lead time of 24 h due to the low RMSE of the raw forecast.

The spatial pattern correlations of the LDA-corrected forecast results also outperformed those of the raw forecasts every day from December 1 to 31, 2018 (Fig. 5). Although the pattern correlation from the LDA-corrected forecast was close to that of the raw forecast for some forecast times, the correction always increased the correlation forecast skill. This indicated that the improvement due to LDA correction was maintained daily and that its performance was stable for the lead time of 24–240 h. The RMSE values from the corrected forecast were lower than those from the raw forecast (Fig. 5). Thus, the LDA-correction method increased the correlation and reduced the RMSE for different lead times and initial forecast times, further establishing the robustness of the LDA-correction method for operational weather forecasts.

The air temperature over the Tibetan Plateau affects surface heat flux and, thus, can affect precipitation over East Asia during summer and winter (Bao et al., 2010; Hsu and Liu, 2003; Wang et al., 2008); therefore, the improvement of T2m forecast accuracy over this region is of great importance. Fig. 6 focuses on the forecast improvement for the Qinghai Tibet Plateau (26° – 39° N, 90° – 104° E). For correlation skill, the improvement percentage from the LDA-correction method in the Qinghai Tibet Plateau was larger than 3% for different lead times. The percentage reached 9% for the lead time of 240 h. The RMSE forecast skill was also improved at different lead times, which indicates the positive role of the LDA-correction method on the forecast accuracy of T2m in

the Qinghai Tibet Plateau.

China's capital city, Beijing, has a very high population density; considering the prevalence of cold-wave events in Beijing in winter, we assessed the forecast skill of mean T2m over the Beijing district (39.5° – 41.0° N, 115.5° – 117.5° E). The forecasts at the lead time of 24 h (1 d) to 240 h (10 d) for 00:00 between December 1 and 31 are shown in Fig. 7. The observation results are represented by a black line (Fig. 7). A cold-wave process can be identified from December 21 to 27, with a cold-wave event occurring on December 22. The ECMWF raw operational forecast results successfully predicted this cooling event. However, there are certain errors between the predicted values of T2m and the observed values. As displayed in Fig. 7, the differences in absolute error from the raw forecast and LDA-correction forecast relative to observation were positive for almost all lead times and initial times, which indicates the reduction of forecast bias from the LDA-correction method. Thus, the correction by the LDA-correction method improved the forecast results from different initial times for different lead times. In particular, in the medium-range forecast stage, the cooling process that occurred on December 23–27 was predicted more accurately than with raw forecast results, such as at lead times of 192, 216, and 240 h. The location of the analogs was restricted by the limited size of the similar sample pool; hence, the range of the improved temperature bias was relatively small. However, this improvement demonstrates that the LDA method has the potential to enhance the forecasting of cold-wave events.

Station observation is independent from ECMWF forecast product for validation; therefore, we also adopted T2m from *in situ* observation to assess the performance of the LDA-correction method. The 410 stations used are shown in Fig. 8. The forecast correction performance in the focused observation stations was evaluated using RMSE and correlation skill. The first two rows of Fig. 9 show the improvement percentage of the correlation skill, and the last two rows correspond to that of RMSE skill in every station. The improvement percentage is the ratio of the difference between the forecast skill of the corrected forecast and that of raw forecast to the raw forecast skill, which is calculated using the following formula: $\text{Improvement percentage} = \frac{\text{skill(corrected)} - \text{skill(raw)}}{\text{skill(raw)}} \times 100\%$, where skill(corrected) is the correlation or RMSE skill between the corrected forecast result and observation and skill(raw) is the correlation or RMSE skill between the raw forecast result and observation in every station. The results shown in Fig. 9 indicate the correlation skill of forecasting improved in most stations upon correction by using the LDA-correction method. The improved stations for correlation skill were concentrated in the north of the focused regions. For the RMSE skill, the improvement in performance was more pronounced. The RMSE for the forecast results corrected by using the LDA-correction method decreased on almost all stations.

Fig. 10 shows the average correlation and RMSE skill over all focused stations. The forecast corrected using the LDA-correction method had higher correlation skill and lower RMSE skill than raw forecast results in the lead times of 24 h to 240 h. The increase in percentage of the correlation skill was from 3% to 18%, and the decrease in percentage of the RMSE skill was from 1.8% to 3.3% in different lead times. Overall, the results showed that, based on the *in situ* observation dataset, the LDA-correction method also improved the accuracy of the ECMWF forecast T2m.

4. Conclusion and discussion

The LDA-correction approach is based on the assumption that the forecast field can be regarded as a small disturbance superimposed on historical analogous states as well as the possibility of incorporating statistical forecasting results into numerical forecasts (Gao et al., 2006; Ren and Chou, 2007; Ren et al., 2009). We applied the LDA-correction method to the ECMWF operational weather forecast. The LDA correction reduced state-dependent error and improved the ECMWF operational T2m product. Over East Asia, the temporal correlation coefficient

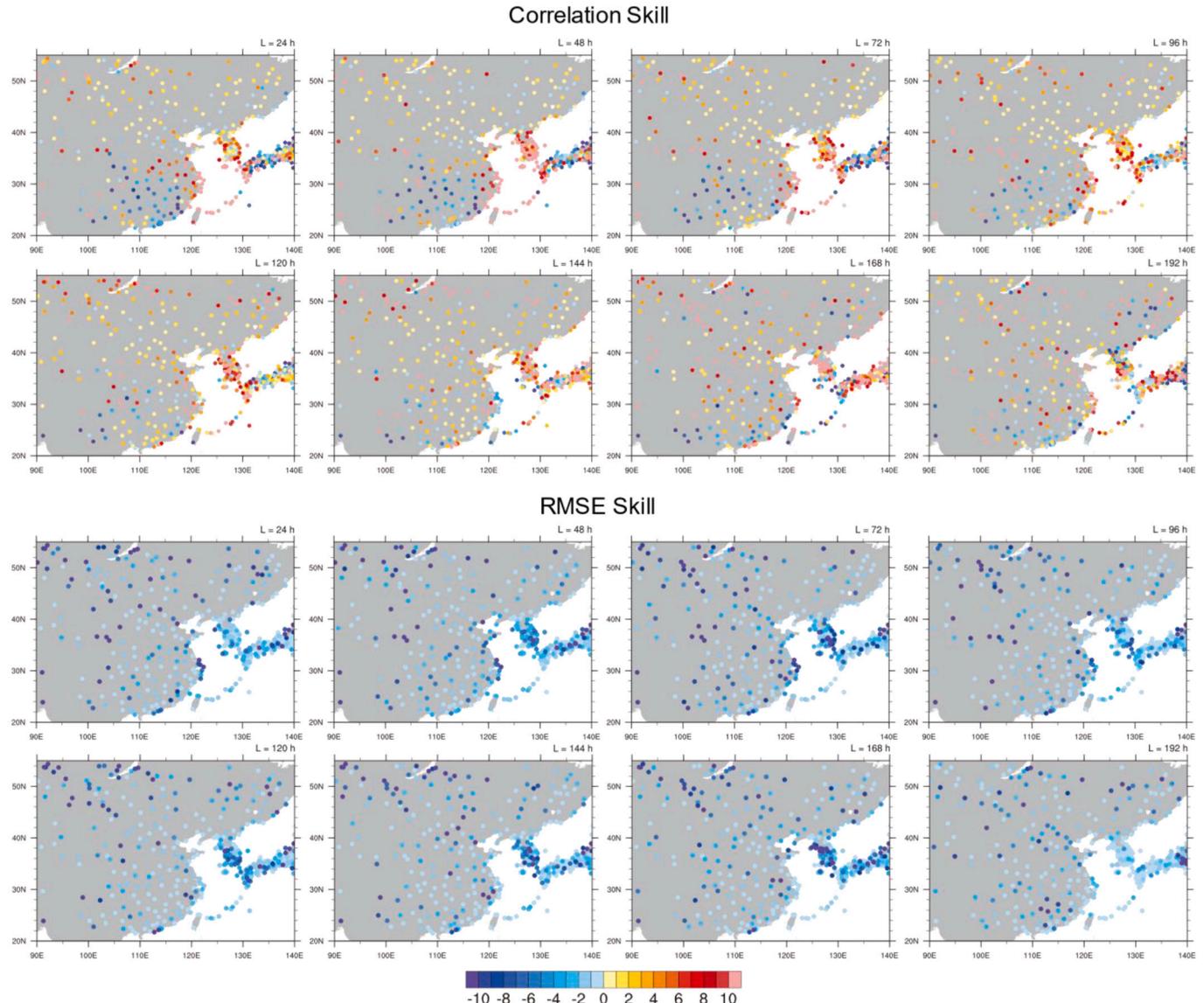


Fig. 9. The improvement percentage of temporal correlation skill (first two rows) and root mean square error (RMSE) skill (last two rows) between raw forecast and observation in the lead times of 24, 48, 72, 96, 120, 144, 168, and 192 h. The improved percentage representing the ratio percentage of the difference between the correlation/RMSE skill of the corrected forecast and that of raw forecast skill.

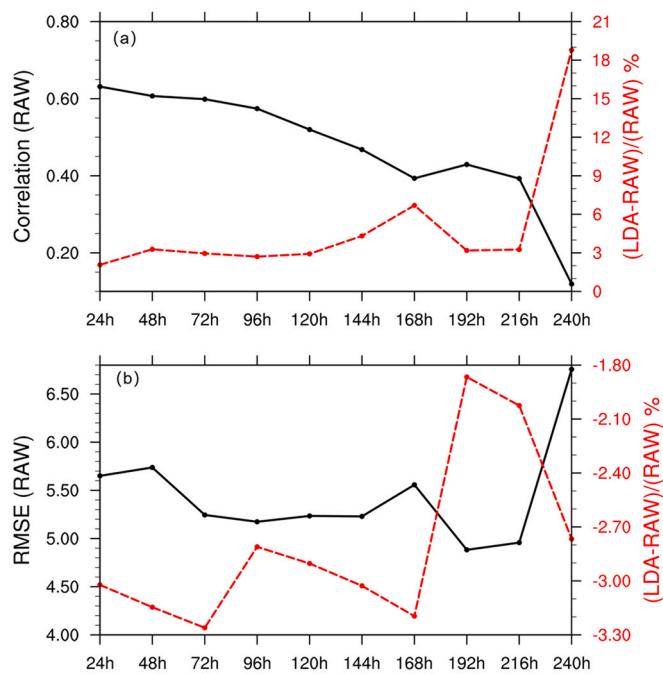


Fig. 10. Spatial average results of temporal correlation (a) and RMSE skill (b) over all the stations in the lead times of 24, 48, 72, 96, 120, 144, 168, 192, and 240 h. The black lines (left vertical coordinates) show the raw forecast skill of the stations from ECMWF, and the red curves (right vertical coordinates) indicate the improvement percentage from using the LDA-correction method relative to raw forecast skill for different lead times. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was higher and RMSE was smaller at different lead times of 24 h to 240 h with LDA correction than those with the raw forecast data. The improvement values vary spatially, which was reflected by the greater improvement in forecasting ability over the Tibetan Plateau. The improvement percentage of the temporal correlation coefficient reached 1%–5%, whereas that of RMSE skill reached 2%–4% for the lead times of 24–240 h. Moreover, the temporal correlation of the Qinghai Tibet Plateau presented a 3%–9% improvement relative to the raw forecast. We also evaluated the forecast correction for a cold-wave process occurring in Beijing and found that the corrected T2m values were closer to the observations than those of the raw forecast values.

In this study, we focused on the period of December 2018. However, we also checked the improvement performance of the LDA-correction method for December 2019 and December 2020. Across different study periods, the LDA-correction method always improved the performance of T2m from ECMWF, although the spatial structure and magnitude of the improvement varied, thereby confirming the robustness of improvement from using the LDA-correction method. In meteorology, interpolation is a commonly used data processing method, as datasets from different sources often have different resolutions (DeGaetano and Belcher, 2007; Kusch and Davy, 2022). ERA5 is often interpolated in the evaluation process and elevation bias is not considered (Pelosi et al., 2020; Soares et al., 2020). In this study, we interpolated T2m of ERA5 from $0.1^\circ \times 0.1^\circ$ to $0.125^\circ \times 0.125^\circ$, which is interpolation from fine to coarse resolution with relatively high accuracy. Therefore, elevation bias resulting from the different resolutions for T2m of the operational weather forecast and ERA5 had minimal

impact on the results of this study.

We also evaluated the improvement in performance of a cold wave that occurred in Beijing. The difference in the focused region influenced the performance improvement. The forecast skill improvement in the Qinghai Tibet Plateau appears large and pronounced (Figs. 2 and 3). Based on the altitude dataset from the ETOPO5 data (<https://www.ngdc.noaa.gov/mgg/global/etopo5.HTML>), we found that altitude affects the improvement in performance from the LDA-correction method. The forecast skill improvement magnitude was higher at higher altitudes and comparatively lesser at lower altitudes (not shown). However, the improvement percentage was higher at lower altitudes and comparatively lesser at higher altitudes, which is attributed to the simulation capability of the forecast model. When raw forecast skill was higher, the predictability in these grids and the improvement proportion from the LDA-correction method were higher, and vice versa (Fig. 11). Therefore, the model's capacity to describe the weather system's mechanism has an impact on its model forecast error and further affects the corrected performance of the LDA-correction method, which reflects the limitations of the analog-corrected method.

In addition to applications in improving model forecast accuracy, the LDA-correction method may be combined with some data assimilation methods to improve the quality of assimilation. The uncertainty of the prior estimate is of great importance in data assimilation and is described by background covariance, which involves the short-range forecast error (Dee and Da Silva, 1998; Feng et al., 2016, 2018). The internationally recognized ensemble Kalman filter uses a set ensemble member to estimate forecast error and calculate the background error covariance (Evensen, 1994). However, the ensemble Kalman filter is limited by its need for substantial computing resources. Based on the analog thought, the forecast error is also estimated using the historical dataset with some analog methods, and analog data assimilation approaches have been proposed (Lguensat et al., 2017; Grooms, 2021). Therefore, the LDA-correction method may be used in data assimilation to optimize the estimation of background covariance in the future.

In this study, we established the applicability of the LDA-correction method in operational weather forecast products. We believe that this study offers a feasible approach for combining historical data and model operational forecast results. In contrast to machine learning, the LDA-correction method considers the impact of the time dimension. Furthermore, the LDA-correction approach requires fewer computing resources and is easily applied to several operational forecast products. The used approach, as well as similar approaches, benefits from historical datasets. Compared to previous studies (Hou et al., 2020, 2021), we used spatially adjacent grids from high-resolution grid data as potential analog pools, for the first time, and verified the applicability of the LDA-correction method in the weather forecast field with small samples.

CRediT authorship contribution statement

Zhaolu Hou: Methodology, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Jianping Li:** Supervision, Conceptualization, Formal analysis, Writing – review & editing. **Lei Wang:** Data curation, Writing – review & editing. **Yazhou Zhang:** Data curation, Writing – review & editing. **Ting Liu:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial

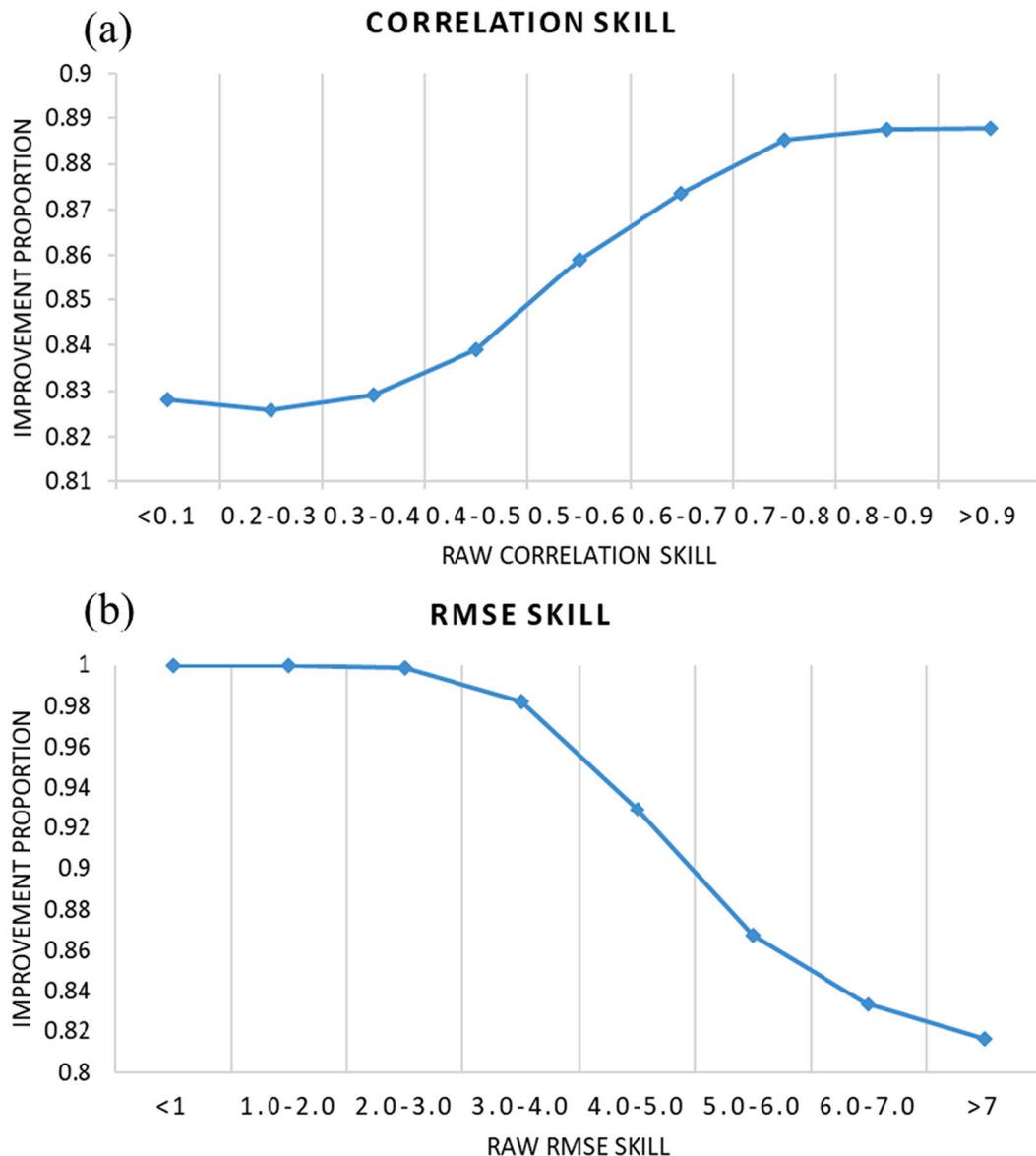


Fig. 11. Proportion of improved grids of correlation (a) and RMSE (b) skill from using the LDA-correction method to all grids in different raw forecast skill intervals. The proportion is average over the lead times of 24, 48, 72, 96, 120, 144, 168, 192, and 240 h. The improved grids represent the grids whose correlation (RMSE) skill of the corrected forecast increases (decreases).

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Shandong Natural Science Foundation Project (ZR2020QD056, ZR2019ZD12), National Natural Science Foundation of China (NSFC) Project (42005049, 42130607), China Postdoctoral Science Foundation (2020M680094), and Fundamental Research Funds for the Central Universities (201962009, 202013031). This manuscript has greatly benefited from the constructive comments offered by the two anonymous reviewers. We acknowledge the support of the Center for High Performance Computing and System Simulation, Qingdao Pilot National Laboratory for Marine Science and Technology. The ERA5 reanalysis data was obtained from <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>.

References

- Bao, Q., Yang, J., Liu, Y.M., Wu, G.X., Wang, B., 2010. Roles of Anomalous Tibetan Plateau Warming on the Severe 2008 Winter storm in Central-Southern China. *Mon. Weather Rev.* 138, 2375–2384.
- Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. *Nature*. 525, 47–55.
- Boukabara, S.A., Krasnopolksy, V., Stewart, J.Q., Maddy, E.S., Shahroudi, N., Hoffman, R.N., 2019. Leveraging modern artificial intelligence for remote sensing and NWP: Benefits and challenges. *Bull. Am. Meteorol. Soc.* 100 (12), ES473–ES491.
- Carter, G.M., Dallavalle, J.P., Glahn, H.R., 1989. Statistical forecasts based on the National Meteorological Center's Numerical Weather Prediction System. *Weather Forecast*. 4, 401–412.
- Chen, H., Chandrasekar, V., Tan, H., Cifelli, R., 2019. Rainfall estimation from ground radar and TRMM precipitation radar using hybrid deep neural networks. *Geophys. Res. Lett.* 46 (17–18), 10669–10678.
- Cho, D., Yoo, C., Im, J., Cha, D.H., 2020. Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. *Earth and Space. Science* 7 (4) (e2019EA000740).
- Clarke, J., 2007. Energy Simulation in Building Design. Routledge.
- Dalcher, A., Kalnay, E., 1987. Error growth and predictability in operational ECMWF forecasts. *Tellus A* 39A, 474–491.

- Danforth, C.M., Kalnay, E., 2008. Impact of online empirical model correction on nonlinear error growth. *Geophys. Res. Lett.* 35 (24), 1–6.
- Dee, D.P., Da Silva, A.M., 1998. Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.* 124 (5545), 269–295.
- DeGaetano, A.T., Belcher, B.N., 2007. Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *J. Appl. Meteorol. Climatol.* 46 (11), 1981–1992.
- DelSole, T., Hou, A.Y., 1999. Empirical correction of a dynamical model. Part I: fundamental issues. *Mon. Weather Rev.* 127, 2533–2545.
- DelSole, T., Zhao, M., Dirmeyer, P.A., Kirtman, B.P., 2008. Empirical Correction of a coupled Land-Atmosphere Model. *Mon. Weather Rev.* 136, 4063–4076.
- Ding, R.Q., Li, J.P., 2007. Nonlinear finite-time Lyapunov exponent and predictability. *Phys. Lett. A* 364, 396–400.
- Ding, R.Q., Li, J.P., Zheng, F., Feng, J., Liu, D.Q., 2016. Estimating the limit of decadal-scale climate predictability using observational data. *Clim. Dyn.* 46, 1563–1580.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* 99 (C5), 10143–10162.
- Eyre, J.R., English, S.J., Forsythe, M., 2020. Assimilation of satellite data in numerical weather prediction. Part I: the early years. *Q. J. R. Meteorol. Soc.* 146, 49–68.
- Feng, J., Ding, R.Q., Li, J.P., Liu, D.Q., 2016. Comparison of nonlinear local Lyapunov vectors with bred vectors, random perturbations and ensemble transform Kalman filter strategies in a barotropic model. *Adv. Atmos. Sci.* 33 (9), 1036–1046.
- Feng, J., Li, J.P., Ding, R.Q., Toth, Z., 2018. Comparison of nonlinear local Lyapunov vectors and bred vectors in estimating the spatial distribution of error growth. *J. Atmos. Sci.* 75 (4), 1073–1087.
- Gao, L., Ren, H.L., Li, J.P., Chou, J.F., 2006. Analogue correction method of errors and its application to numerical weather prediction. *Chin. Phys.* 15, 882–889.
- Glahn, H.R., Lowry, D.A., 1972. The use of Model output Statistics (MOS) in Objective Weather forecasting. *J. Appl. Meteorol.* 11, 1203–1211.
- Grooms, I., 2021. Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders. *Q. J. R. Meteorol. Soc.* 147 (734), 139–149.
- Haiden, T., Janousek, M., Bidlot, J., Buizza, R., Ferranti, L., Prates, F., Vitart, F., 2018. Evaluation of ECMWF forecasts. In: including the 2018 upgrade. European Centre for Medium Range Weather Forecasts.
- Hou, Z.L., Zuo, B., Zhang, S.Q., Huang, F., Ding, R.Q., Duan, W.S., Li, J.P., 2020. Model forecast error correction based on the local dynamical analog method: an example application to the ENSO forecast by an intermediate coupled model. *Geophys. Res. Lett.* 47 (e2020GL088986).
- Hou, Z.L., Li, J.P., Zuo, B., 2021. Correction of Monthly SST forecasts in CFSv2 using the Local Dynamical Analog Method. *Weather Forecast.* 36, 843–858.
- Hsu, H.H., Liu, X., 2003. Relationship between the Tibetan plateau heating and east Asian summer monsoon rainfall. *Geophys. Res. Lett.* 30 (20).
- Huld, T., Gottschalg, R., Beyer, H.G., Topic, M., 2010. Mapping the performance of PV modules: effects of module type and data averaging. *Sol. Energy* 84, 324–338.
- Ji, L., Wang, Z., Chen, M., Fan, S., Wang, Y., Shen, Z., 2019. How Much Can AI Techniques Improve Surface Air Temperature Forecast?—A Report from AI Challenger 2018 Global Weather Forecast Contest.
- Kam, J., Kim, S., Roundy, J.K., 2021. Did a skillful prediction of near-surface temperatures help or hinder forecasting of the 2012 US drought? *Environ. Res. Lett.* 16, 034044.
- Kim, H.M., Webster, P.J., Curry, J.A., 2012. Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Clim. Dyn.* 39, 2957–2973.
- Koehl, M., Heck, M., Wiesmeier, S., Wirth, J., 2011. Modeling of the nominal operating cell temperature based on outdoor weathering. *Sol. Energy Mater. Sol. Cells* 95, 1638–1646.
- Kusch, E., Davy, R., 2022. KrigR—A tool for downloading and statistically downscaling climate reanalysis data. *Environ. Res. Lett.* 17.
- Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., Fablet, R., 2017. The analog data assimilation. *Mon. Weather Rev.* 145 (10), 4093–4107.
- Li, J.P., Ding, R.Q., 2011. Temporal-Spatial distribution of Atmospheric Predictability Limit by Local Dynamical Analogs. *Mon. Weather Rev.* 139, 3265–3283.
- Li, J.P., Ding, R.Q., 2013. Temporal-spatial distribution of the predictability limit of monthly sea surface temperature in the global oceans. *Int. J. Climatol.* 33, 1936–1947.
- Li, J.P., Ding, R.Q., 2015. Seasonal and interannual weather prediction. In: North, G., Pyle, J., Zhang, F. (Eds.), *Encyclopedia of Atmospheric Sciences*, , 2nd edition vol. 6. Academic Press and Elsevier, pp. 303–312.
- Li, J.P., Wang, S., 2008. Some mathematical and numerical issues in geophysical fluid dynamics and climate dynamics. *Communications in Computational Physics* 3, 759–793.
- Lorenc, A.C., 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* 112, 1177–1194.
- Lynch, P., 2006. *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge University Press.
- Marzban, C., 2003. Neural networks for postprocessing model output: ARPS. *Mon. Weather Rev.* 131 (6), 1103–1111.
- McTaggart-Cowan, R., Vaillancourt, P.A., Zadra, A., Chamberland, S., Charron, M., Corvec, S., Milbrandt, J.A., Paquin-Ricard, D., Patoine, A., Roch, M., Separovic, L., Yang, J., 2019. Modernization of atmospheric physics parameterization in Canadian NWP. *Journal of Advances in Modeling Earth Systems.* 11, 3593–3635.
- NOAA National Centers for Environmental Information, 2001. Global Surface Hourly Integrated Surface Dataset. NOAA National Centers for Environmental Information. <https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ndcc:C00532.html>.
- Pelosi, A., Terribile, F., D'Urso, G., Chirico, G.B., 2020. Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water* 12 (6), 1669.
- Peng, T., Zhi, X., Ji, Y., Ji, L., Tian, Y., 2020. Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning post-processing methods. *Atmosphere* 11 (8), 823.
- Rasp, S., Lerch, S., 2018. Neural networks for postprocessing ensemble weather forecasts. *Mon. Weather Rev.* 146 (11), 3885–3900.
- Ren, H.L., Chou, J.F., 2007. Strategy and methodology of dynamical analogue prediction. *Sci. China Ser. D Earth Sci.* 50, 1589–1599.
- Ren, H.L., Chou, J.F., Huang, J.P., Zhang, P.Q., 2009. Theoretical basis and application of an analogue-dynamical model in the Lorenz system. *Adv. Atmos. Sci.* 26, 67–77.
- Shen, X., Wang, J., Li, Z., Chen, D., Gong, J., 2020. Research and Operational Development of Numerical Weather Prediction in China. *Journal of Meteorological Research.* 34, 675–698.
- Soares, P.M., Lima, D.C., Nogueira, M., 2020. Global offshore wind energy resources using the new ERA-5 reanalysis. *Environ. Res. Lett.* 15 (10), 1040a2.
- Sweeney, C.P., Lynch, P., Nolan, P., 2013. Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. *Meteorol. Appl.* 20, 32–40.
- Thomas, G., Mahfouf, J.-F., Montmerle, T., 2020. Toward a variational assimilation of polarimetric radar observations in a convective-scale numerical weather prediction (NWP) model. *Atmospheric Measurement Techniques.* 13, 2279–2298.
- Tian, B.Q., Fan, K., Yang, H.Q., 2018. East Asian winter monsoon forecasting schemes based on the NCEP's climate forecast system. *Clim. Dyn.* 51, 2793–2805.
- Vannitsem, S., Bremnes, J.B., Demaeyer, J., Evans, G.R., Flowerdew, J., Hemri, S., et al., 2021. Statistical Postprocessing for Weather forecasts: Review, challenges, and Avenues in a big Data World. *Bull. Am. Meteorol. Soc.* 102 (3), E681–E699.
- Vislocky, R.L., Young, G.S., 1989. The use of perfect prog forecasts to improve model output statistics forecasts of precipitation probability. *Weather Forecast.* 4 (2), 202–209.
- Wang, B., Bao, Q., Hoskins, B., Wu, G.X., Liu, Y.M., 2008. Tibetan plateau warming and precipitation changes in East Asia. *Geophys. Res. Lett.* 35.
- Zhu, Z.W., Li, T., 2017. Statistical extended-range forecast of winter surface air temperature and extremely cold days over China. *Q. J. R. Meteorol. Soc.* 143, 1528–1538.