Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2024-020
J-STAGE Advance published date: March 27th, 2024
The final manuscript after publication will replace the preliminary version at the above DOI once it is available.

Development of a Temperature Prediction Method Combining Deep Neural Networks and a Kalman Filter

Takuya INOUE[1]

Numerical Prediction Development Center
Japan Meteorological Agency, Tsukuba, Japan

Tsuyoshi Thomas SEKIYAMA

Meteorological Research Institute
Japan Meteorological Agency, Tsukuba, Japan

and

Atsushi KUDO

Numerical Prediction Division
Japan Meteorological Agency, Tokyo, Japan

September 19, 2023

----

1) Corresponding author: Takuya INOUE, Numerical Prediction Development Center, Japan Meteorological Agency, 1-2, Nagamine, Tsukuba, Ibaraki 305-0502 JAPAN
Email: tak_inoue@met.kishou.go.jp

Abstract

Numerical weather forecast models have biases caused by insufficient grid resolution and incomplete physical processes, especially near the land surface. Therefore, the Japan Meteorological Agency (JMA) has been operationally post-processing the forecast model outputs to correct biases. The operational post-processing method uses a Kalman filter (KF) algorithm for surface temperature prediction. Recent reports have shown that deep convolutional neural networks (CNNs) outperform the JMA operational method in correcting temperature forecast biases. This study combined the CNN-based bias correction scheme with the JMA operational KF algorithm. We expected that the combination of CNNs and a KF would improve the post-processing performance, as the CNNs modify large horizontal structures, and then, the KF corrects minor spatiotemporal deviations. As expected, we confirmed that the combination outperformed both CNNs and the KF alone. This study demonstrated the advantages of the new method in correcting coastal fronts, heat waves, and radiative cooling biases.

Keywords deep convolutional neural network, statistical post-processing, temperature forecast, Kalman filter, fine-tuning

1. Introduction

Temperature is an element of weather that has a large impact on daily life as well as social, agricultural, and economic activities. Numerical weather prediction (NWP) is commonly used for forecasting temperatures. However, NWP models have biases due to limited horizontal grid resolution and imperfections in physical processes. Thus, the Japan Meteorological Agency (JMA) has been operationally post-processing the NWP model outputs to correct these biases. This post-processing is called guidance (Klein and Glahn 1974; Zurndorfer et al. 1979) or model output statistics (MOS; Glahn and Lowry 1972). The JMA provides temperature guidance products to support forecasters in short-range surface temperature forecasts (JMA 2023a). Furthermore, the JMA has improved temperature guidance forecasts to prevent heatstroke from extreme temperatures or crop damage from low temperatures. The JMA also aims to improve transportation safety by improving snowfall forecasts that use temperature guidance forecasts (Furuichi and Matsuzawa 2009).

At present, the JMA has two types of temperature guidance systems in operation: a point-like temperature guidance system and a gridded temperature guidance system (Sannohe 2018). The JMA started operating a point-like temperature guidance system in 1979 (JMA 1986), and a Kalman filter (KF) was introduced into the algorithm in 1996 (Segami et al. 1995). The point-like temperature guidance system forecasts 1.5 m of temperature at each meteorological station. The equations were adjusted successively at more than 900 Japanese stations in the Automated Meteorological Data Acquisition System (AMeDAS; JMA 2023b). The explanatory variables are the NWP outputs around the stations, and the objective variable is the temperature difference between the NWP outputs and observations at the stations. By statistically correcting NWP model biases, a temperature guidance system can reduce forecast errors in NWP models. However, the operational guidance system of the JMA cannot correct horizontal positioning errors, such as positional errors in coastal fronts (Takada 2018a), because it uses only explanatory variables around the stations.

JMA's temperature guidance employs an online learning technique with a KF that sequentially evolves the coefficients of the prediction equations based on the latest observations. Online learning has four advantages: it can follow seasonal changes in NWP biases, NWP model updates (Takada 2018b), and changes in the environment due to observatory relocation (Takada 2018c), and it can adapt to newly established observatories without a long-term dataset. The most important advantage is that online learning has the ability to respond to NWP model updates. NWP models are regularly updated to increase performance (Wilson and Vallée 2002). As NWP models change, the biases in NWP models change, meaning that post-processing must be reconfigured with a new dataset. Online learning with the KF can accommodate these changes. The second is that it can respond to changes in the surrounding areas of stations. AMeDAS stations are relocated if their environmental conditions change. When a station has relocated, the characteristics at that location often change significantly (Miura and Ohashi 2017). The guidance system can adapt to new locations through online learning without requiring a long-term observational dataset.

The other temperature guidance forecasts, i.e., gridded temperature guidance forecasts, are created from point-like temperature guidance forecasts and gridded temperature predictions of the NWP models by weighted averaging based on distance and topography (Kuroki 2017). Because the JMA's operational gridded temperature guidance system links to the point-like temperature guidance system, there is consistency between point-like and gridded temperature guidance forecasts.

National weather agencies utilize post-processing algorithms for forecasting temperatures. The National Weather Service uses multiple linear regressions (MLRs) to generate both point and grid temperature guidance forecasts. They objectively analyze guidance forecasts with elevation corrections to produce gridded forecasts of weather elements, such as temperature, clouds, and snow amount (Glahn et al. 2009). The gridded guidance forecasts are spatially consistent predictions that are provided for forecasters. The Met Office employs KF for point-like temperature (Met Office 2015) and physically based corrections for height differences between the terrain in the NWP models and the actual topography for gridded temperature (Sheridan et al. 2010). Météo-France provides point-like temperature predictions using MLR, KF and random forest (Météo-France 2015, Météo-France 2020). Deutscher Wetterdienst used MLR for point-like temperature forecasts (Veira et al. 2017). To our knowledge, no national weather agency currently uses deep learning methods for temperature forecasting post-processing.

Recently, several studies have been conducted on temperature predictions via deep-learning methods. To our knowledge, studies have yet to combine gridded and point-like forecasts. Dongjin et al. (2022) compared several machine learning and deep learning methods and showed that convolutional neural networks (CNNs) were effective at post-processing next-day maximum temperatures. They reported that CNNs performed well among the other post-processing models by using spatial information surrounding stations; however, they did not refer to the relocation of stations. In general, it is impossible to train networks until sufficient observation data are stored at a new site after relocation. In the study of gridded temperature forecasting, Bing et al. (2022) verified convolutional long short-term memory (ConvLSTM; Shi et al. 2015) models as a forecasting method for timeseries gridded temperatures. They applied them to create hourly forecasts of the 2-m temperature for the subsequent 12 h over Europe. Although these methods did not reach the capabilities of current NWP models, they demonstrated that deep neural networks may achieve forecast quality beyond the nowcasting range in a data-driven way. Kudo (2022) studied gridded forecasts of 1.5-m temperature using CNNs. They reported that the CNN has the ability to correct the horizontal position bias in temperatures in NWP models. Their “DNN-based gridded temperature predictions” surpassed the JMA's operational gridded temperature guidance forecast by approximately 0.25°C in terms of the root mean square error (RMSE). Furthermore, their study showed that the CNN corrects NWP model biases, such as positional errors of coastal fronts and extreme temperatures, which are difficult to predict in the operational guidance forecast of the JMA. However, their study did not focus on point-like predictions; therefore, the performance at each station is uncertain.

The present study combined the bias corrections of CNNs and the KF to produce point-like temperature predictions. Since CNNs can correct the large horizontal structure of NWP models and KFs can correct small spatiotemporal errors, we expect that the combination of each method will improve post-processing performance. In addition, the method could adapt the relocations of stations and NWP model updates through online learning with the KF.

2. Data and Methodology

2.1 Meteorological data

Following a previous study (Kudo 2022), the present study used the operational mesoscale nonhydrostatic regional model (MSM; JMA 2023c) outputs of the JMA for explanatory variables with a 5-km horizontal resolution and a three-hour interval. The dataset period ranged from 00 UTC on October 8, 2010, to 21 UTC on December 31, 2021, with the MSM forecasts initialized at 00, 03, 06, 09, 12, 15, 18, and 21 UTC. For training the CNNs, we used only 15-hour predictions from each initial time, as in Kudo (2022). However, the CNN inference forecast range was 3 to 39 hours at 3-hour intervals to clarify the performance of the CNNs.

The objective variable was the 1.5-m temperature extracted from the operational estimated weather distribution products of the JMA (Wakayama et al. 2020). The products are 1-km grid data of hourly temperature, weather category, and sunshine duration over land in Japan. The temperature is estimated from observed temperatures and the gridded climatological normal temperature calculated by the JMA. The gridded climatological normal is estimated from gridded data of climatological normal from the most recent 30 years at each observatory. It is calculated by MLRs based on the statistical relationship between normal and topographic/urban factors. The estimated temperatures are generated by interpolating observations with the gridded climatological normal. Therefore, the estimated temperatures are expected to be close to reality, even in areas where there are no observation sites. The cross-validation of the estimated temperature showed that the bias was approximately 0°C and that the RMSE was approximately 1°C (JMA 2016).

We averaged the estimated temperatures in 5-km grids following the MSM grids. The estimated surface temperature (EST) in the 5-km grid dataset served as the target or ground truth for the gridded prediction, i.e., the observational temperature distribution. The dataset covered the same period as that of the MSM forecast.

2.2 Structure of the neural networks

Fig. 1 shows the CNN model used in the present study, which is the same as the encoder– Fig. 1

Table 1

decoder-based deep CNNs proposed in Kudo (2022). The CNN model consisted of 2-dimensional convolution, max-pooling, and fully connected layers with sigmoid or ReLU (Nair and Hinton 2010) activation functions and batch normalization. Table 1 describes the parameters used in the model. The network input seven types of variables and output a 1.5-m temperature with 128 × 128 grid points at 5 km intervals. The seven types of input variables were surface temperature; temperatures at 975, 925, and 850 hPa; mean sea level pressure (MSLP); and surface wind components U and V derived from the MSM. These explanatory variables are empirically selected using the training and validation datasets by Kudo (2022). The surface temperature is a physical quantity in the MSM outputs that has the characteristics closest to the objective variable. It is considered highly correlated with EST. Temperatures at 975, 925, and 850 hPa are expected to represent the impact of the atmospheric boundary layer on surface temperatures through the vertical turbulent transport of heat. The locations of low-pressure systems and fronts can be estimated from MSLP and surface wind data, providing overviews of the synoptic situation. For the JMA's operational point-like temperature guidance system, surface temperature and wind are used as explanatory variables (Sannohe 2018). The input variables were standardized with each input channel's maximum and minimum values ranging between 0 and 1. After encoding and decoding, the output variables were inversely transformed. The CNN model was trained with the EST for each forecast lead time using the mean square error loss function with the Adam optimizer (Kingma and Ba 2015). The input and target

Table 2

datasets were divided into three parts—training, validation, and test periods—as shown in Table 2. The validation dataset was used only for hyperparameter adjustment. The test dataset was used to verify the prediction accuracy of the CNN model.

2.3 Prediction procedure

2.3.1 CNN model prediction

Fig. 2

This study defined six areas (jp01, jp02, jp03, jp04, jp05, and jp06) as target domains to cover most of Japan, as shown in Fig. 2. Each domain had 128 × 128 grid points to cover the whole area of Japan's second-largest island, Hokkaido (jp01). While Kudo (2022) implemented CNN model prediction with a size of 64 × 64 grid points to cover the area around Tokyo, we doubled the size and targeted nearly all of the Japanese archipelago. We trained the CNN model at each target domain separately to reduce the consumption of GPU memory and calculation time. In addition, it was appropriate to train the networks separately in domains because each domain had different meteorological and climatological properties with different land-to-sea ratios.

The study introduced a fine-tuning procedure, which retrains the networks using the data immediately preceding the validation period, from January 1 to December 31, 2019, to correct for long-term trends in the NWP models. One of the advantages of applying fine-tuning in a short training period is that it takes less time than reconstructing the network in a long training period. By applying fine-tuning, the network can be trained on NWP models without using a long-term training dataset. This approach is favorable for operational systems with frequent NWP model updates.

2.3.2 Post-processing with a Kalman filter

The JMA's operational point-like temperature guidance system uses a KF to predict surface temperatures at each observatory. In addition, in-situ observations and NWP outputs are used as input data. The NWP outputs are interpolated from the surrounding grids to the forecast points. In the guidance system, the predictand (i.e., the target of forecasting) is defined as the temperature difference between the NWP outputs and observations. The prediction equation is represented by a linear combination of predictors and coefficients as follows (JMA 2023a):

$$y_{\tau+1} = \boldsymbol{c}_{\boldsymbol{\tau}+\mathbf{1}}\boldsymbol{X}_{\boldsymbol{\tau}+\mathbf{1}}$$

where τ represents the sequence number of NWP initial times, $y_{\tau+1}$ represents the predictand, $\boldsymbol{c}_{\boldsymbol{\tau}+\mathbf{1}}$ represents the predictors ($1 \times \mathrm{n}$ matrix), and $\boldsymbol{X}_{\boldsymbol{\tau}+\mathbf{1}}$ represents the coefficients ($\mathrm{n} \times 1$ matrix). The coefficients $\boldsymbol{X}_{\boldsymbol{\tau}+\mathbf{1}}$ are determined from both the previous estimate $\boldsymbol{X}_{\boldsymbol{\tau}}$ and the forecast error to minimize the diagonal sum of the error covariance matrix. This indicates that the coefficients are optimized at each initial time based on the difference between the previous forecast and the observations. As a result, the system with KFs has the flexibility to follow seasonal changes, NWP model updates, and changes due to observatory relocation.

The purpose of this study is to develop a post-processing system for DNN-based gridded forecasts with a KF. Hereafter, we call this the "DNN-based point-like temperature guidance forecast" (DNN-KF). The KF algorithm to be introduced in DNN-KF is the same as that of JMA's operational point-like temperature guidance forecast.

The DNN-KF generated temperature predictions in the following two steps. First, the trained CNN model generated gridded temperature forecasts. Second, online learning with the KF was applied for each station. In the first step, the CNN model corrected large-scale structural biases, while in the second step, the KF model corrected point- and season-dependent spatiotemporal biases. By constructing a dual-processing system, we expected to improve the forecast accuracy by removing both large- and local-scale biases.

As shown in Table 2, we set the training and test periods of the KF so as not to overlap with the training, fine-tuning, and validation periods of the CNN model. The initial coefficients were copied from the operational guidance system on December 31, 2019.

2.4 Verification method

The verification metric in the study is the RMSE, which is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{n=1}^{N}(F_{nt} - O_{nt})^2},$$

where T and N denote the numbers of time slices and observation points, respectively. F_{nt} and O_{nt} denote the predicted and observed temperatures at point n and time t, respectively.

The relative improvement, or skill score (Wilks 2011), is defined as a reduction in the RMSE normalized by the RMSE for a reference forecast,

$$\text{relative improvement} \equiv \frac{\mathrm{RMSE}_{ref} - \mathrm{RMSE}_{tgt}}{\mathrm{RMSE}_{ref}} \times 100,$$

where $RMSE_{ref}$ is the RMSE for a reference forecast and $RMSE_{tgt}$ is the RMSE for a targeted forecast.

We compared the DNN-KF with the predictions of MSM, operational point-like/gridded temperature guidance (point-like/gridded MSM-KF), and "DNN-based gridded temperature prediction (DNN)." Both the point-like MSM-KF and DNN-KF predict temperatures at observation sites using KF. The point-like MSM-KF/DNN-KF is derived from the MSM/DNN along with in-situ observations. The point-like predictions are verified by calculating the RMSE at the observation sites, while the gridded predictions are verified by linearly interpolating the predictions to the observation sites. The MSM/DNN verified at each observatory is denoted as "interpolated MSM/DNN."

3. Results and Discussion

3.1 Averaged scores

Figure 3 shows the monthly averaged RMSEs of each forecast for the test period. The green, blue, brown, and red lines indicate the interpolated MSM, the point-like MSM-KF, the interpolated DNN, and the DNN-KF, respectively. As shown in the figure, the DNN-KF outperforms the other predictions throughout the period. Fig. 3

Figure 4 shows the average RMSEs of each forecast classified by forecast lead times for the one-year test period from January 1 to December 31, 2021. The results indicate that the Fig. 4

DNN-KF is superior to the other methods in terms of the forecasting lead times.

Fig. 5

Figure 5a shows the relative improvement in the interpolated DNN over the interpolated MSM, and Fig. 5b shows that the improvement in the DNN-KF over the DNN. The red points represent improvement, and the blue points represent deterioration. The RMSEs improved at most stations. These results revealed that the combination outperformed the CNNs or the KF alone. Kudo (2022) also showed that the DNN is more accurate than the MSM by training the DNN with the MSM outputs and EST including in-situ observations. The higher accuracy of the DNN-KF over the DNN is explained by the fact that the KF learns the error characteristics of the locations and has no interpolation errors. The operational point-like MSM-KF also uses in-situ observations through online learning and has no interpolation error. However, the DNN-KF has a higher accuracy than the MSM-KF as shown in Fig. 4, at least on an annual average basis, partly because the DNN is more accurate than the MSM as input data.

3.2 Case studies

3.2.1 Coastal front positioning error

On December 29, 2021, a sharp temperature change caused by a coastal front occurred in the Kanto (jp03) region. Figure 6a shows the observational temperature distribution. The coastal front was close to the estimated 10°C isothermal line in the southern part of the region.

Fig. 6

Figures 6b, 6c, and 6d show each 5-km gridded temperature prediction difference in the

MSM, gridded MSM-KF, and DNN from the EST, respectively. The MSM and gridded MSM-KF predicted the coastal front further north than the actual position. In contrast, the DNN predicted the position of the 10°C isothermal line as being close to the actual position. Consequently, the DNN substantially reduced the errors at Nerima (marked by the cross). Figure 7 shows the time series of observed and predicted temperatures initialized at 21 LST or 12 UTC on December 28, 2021, at Nerima. The point-like MSM-KF predicted temperatures higher than the observations (OBS), while the DNN-KF predicted temperatures closer to the OBS than did the interpolated MSM and MSM-KF.

Several previous studies reported that the MSM has systematic errors in forecasting coastal fronts north of their actual position (Hara 2014; Kawano et al. 2019). Suzuki et al. (2021) used the MSM to conduct sensitivity experiments. These authors discovered that differences in topography between reality and NWP models can cause positional errors. The authors insist that the positional error is a bias that statistical methods can remove. However, some biases cannot be adequately removed by the point-like MSM-KF (Sannohe 2018). One of the possible reasons is that the point-like MSM-KF only uses explanatory variables from the grids surrounding the target point. Conversely, the CNN model uses explanatory variables from the entire target area so that the DNN can correct positional errors associated with coastal fronts.

3.2.2 Heat wave

Fig. 8

On July 1, 2022, the maximum temperature exceeded 35°C in the inland area of the Kanto region (Fig. 8a). The temperatures of the MSM and the gridded MSM-KF were lower than those of the EST. In contrast, the DNN agreed with the EST, especially in the heat wave area. Notably, the MSM has a negative bias in predicting daytime surface temperatures in summer (Hara and Kurahashi 2017; Kusabiraki and Moriyasu 2013). Kusabiraki (2020) indicated that the large negative bias in the MSM was due to excessive upper-level cloud coverage and subsequent insufficient downward shortwave radiation at the surface. To eliminate these issues, cloud microphysical processes improved in 2020 (JMA 2021). In 2022, evapotranspiration processes improved to further reduce the negative bias (JMA 2022). However, negative bias was not completely eliminated. The DNN could efficiently correct the negative bias in this case.

Fig. 9

Figure 9 shows the time series of observed and predicted temperatures initialized at 15 LST or 06 UTC on June 30, 2022, at Tokyo (shown in Fig. 8). Temperatures on July 1, 2022, predicted by the interpolated MSM and point-like MSM-KF were lower than that of OBS. The interpolated DNN adjusted the interpolated MSM prediction moderately in the morning but excessively in the afternoon, causing the interpolated DNN to be much greater than the OBS at 15 and 18 LST. The training data for the DNN included only the period of 2012–2019, which was before the reduction in the MSM negative bias. This result is probably the reason for the excessive adjustment of the DNN in the afternoon, as the MSM prediction in 2022 was performed by the bias-reduced version. However, the DNN-KF successfully corrected the excessive adjustment of the DNN. Since the online learning of the DNN-KF was continuously performed from 2020 to the present (June 30, 2022), the DNN-KF learned the tendency for excessive DNN adjustment.

Figure 10 shows the interannual changes in the ME and the RMSE at each observatory in the Kanto region at 15 LST from 2020 to 2022 in summer. In 2020 and 2021, the negative biases of the interpolated MSM were large, and those of the interpolated DNN and the DNN-KF were close to zero. In 2022, the negative bias of the interpolated MSM was reduced, and the interpolated DNN had a positive bias, but the bias of the DNN-KF remained close to zero. The RMSE of the DNN-KF was also smaller than that of the interpolated MSM and DNN. This result demonstrated that the combination of the two methods, i.e., the DNN and KF, resulted in better forecasts, indicating the robustness of the DNN-KF to minor changes in forecast models.

3.2.3 Low temperature caused by radiative cooling

The MSM and MSM-KF exhibit poor performance in predicting low temperatures caused by radiative cooling (Sannohe 2018), as temperature decreases due to radiative cooling vary greatly depending on weather conditions, such as clouds and wind, and it is difficult to accurately predict these factors with current NWP models. However, according to Kudo (2022), CNNs can predict such low temperature cases because they use surface and lower troposphere temperatures along with MSLP and wind components as predictors. This is because the bias in the MSM surface temperature becomes larger when the temperature lapse rate in the lower troposphere is close to the dry adiabatic lapse rate, such as at a time of radiative cooling.

Fig.

In the early morning on November 16, 2021, the clear sky enhanced radiative cooling, inducing low temperatures in eastern Hokkaido (jp01), as shown in Fig. 11a (at 15 LST on November 16 or 21 UTC on November 15, 2021). The EST indicates a temperature of less than -6°C in the plain of eastern Hokkaido around Shibecha (marked by the cross). Figures 11b, 11c, and 11d show the temperature differences initialized at 21 LST on November 14, 2021. The MSM and gridded MSM-KF temperatures were higher than those of the EST in eastern Hokkaido. Figure 11d shows that the DNN was closer to the EST than the other DNNs were. The CNN model could correct the low temperature bias induced by radiative cooling.

Fig.

Figure 12 shows the time series of observed and predicted temperatures initialized at 21 LST on November 14, 2021, at Shibecha. The MSM predicted temperatures higher than the OBS. The MSM-KF roughly corrected the interpolated MSM bias. The interpolated DNN was also higher than the OBS, although it was better than the interpolated MSM prediction. The DNN-KF was the most accurate prediction, as it successfully corrected the temperature bias.

These results showed that the DNN outperformed the MSM in terms of the low temperatures caused by radiative cooling. The DNN-KF improved the DNN. However, these CNN-based schemes failed to correct the temperature bias outside the eastern part of Hokkaido, where the CNN-based error correction did not work effectively.

4. Conclusion

We propose a new method for point-like temperature predictions that is more accurate than the point-like MSM-KF, the JMA's operational point-like temperature guidance forecast. To generate point-like forecasts from gridded predictions, we adopted a KF. As a result, the new method outperformed the point-like MSM-KF. The DNN-KF outperformed the MSM-KF in terms of the 6-h to 39-h forecast lead times throughout the test period. Furthermore, the DNN successfully corrected NWP model biases, such as coastal front positioning errors and extreme temperatures, which are difficult to correct by the MSM-KF. Our case study revealed that the KF was capable of correcting forecast errors of the DNN caused by NWP model updates through online learning. This study showed that the combination of DNNs and a KF can generate more accurate temperature predictions at each observatory. Our method has the ability to predict extreme low temperatures in a radiative cooling case where operational guidance could not. However, it is still difficult to adequately predict radiative cooling cases, so we need to identify the conditions under which our method does not work well.

We further improve the CNNs to increase the prediction accuracy by our proposed method of combining CNNs and a KF. We intend to find a more appropriate set of hyperparameters and input variables for training CNNs. We would also like to find more suitable network constructions by trying other models, such as U-Net and ResNet. The inputs to the CNNs were the mesoscale model results, which are among the NWP products of the JMA; however, replacing the input with a global or local scale model is a candidate for future experiments. We also consider the use of multiple NWP models as inputs to CNNs rather than as single NWP models.

Inputting NWP outputs into deep learning models such as CNNs is already gaining momentum in this area. Our method corrects NWP outputs with not only CNNs but also KFs that the JMA conventionally uses for post-processing. Many national weather agencies use conventional machine learning methods, such as KFs, MLRs, and neural networks, for operational post-processing of NWP outputs. It will be interesting to see if this combination of deep learning methods and their operating machine learning methods will also be effective at post-processing for their NWP outputs. Operational NWP models are updated regularly in general. We have shown the ability of the DNN-KF following changes in NWP biases through online learning with KFs. This method can be applied to outputs from other NWP models. We expect that our method will lead to an improvement in the operational post-processing of NWP outputs.

Data Availability Statement

The model source codes used in this study are available subject to a license agreement with the JMA headquarters. The datasets of the mesoscale model outputs of the JMA were operationally provided via the Japan Meteorological Business Support Center (http://www.jmbsc.or.jp/en/index-e.html) and are freely available for research purposes.

Acknowledgments

This work was supported by the Japanese Society for the Promotion of Sciences (JSPS) KAKENHI (Grant Number JP21H03593).

References

Bing, G., M. Langguth, Y. Ji, A. Mozaffari, S. Stadtler, K. Mache, and M. G. Schultz, 2022: Temperature forecasting by deep learning methods. Geosci. Model Dev., 15, 8931–8956, doi:10.5194/gmd-15-8931-2022.

Dongjin, C., Y. Cheolhee, S. Bokyung, I. Jungho, Y. Donghyuck, and C. Dong-Hyun, 2022: A novel ensemble learning for post-processing of NWP Model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches. Wea. Climate Extremes, 35, doi:10.1016/j.wace.2022.100410.

Furuichi, Y., and N. Matsuzawa 2009: Snowfall amount guidance. Textbook for Numerical Weather Prediction. 47, Japan Meteorological Agency, 27-38 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwptext/47/chapter2.pdf.]

Glahn, B., K. Gilbert, R. Cosgrove, D. P. Ruth, and K. Sheets, 2009: The gridding of MOS.

Wea. Forecasting, 24, 520-529, doi:10.1175/2008WAF2007080.1.

Glahn, H. R., D. A. Lowry, 1972: The Use of Model Output Statistics (MOS) in Objective Weather Forecasting, J. Appl. Meteor. Climatol., 11, 1203-1211, doi:10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

Hara, T., 2014: Studies on recent remarkable cases. Textbook for Numerical Weather Prediction. 47, Japan Meteorological Agency, 118-144 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwptext/47/chapter4.pdf.]

Hara, T., and H. Kurahashi, 2017: Changes in the characteristics of meso-scale numerical prediction system. Textbook for Numerical Weather Prediction. 50, Japan Meteorological Agency, 48-55 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwptext/50/chapter2.pdf.]

Japan Meteorological Agency, 1986: Progress of guidance in Japan. Textbook for Weather Prediction Technique, 35, Japan Meteorological Agency, 19-20 (in Japanese)Japan Meteorological Agency, 2021: Development results. Numerical Prediction Development Center Annual Report 2020. Japan Meteorological Agency, 15-76 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/npdc/r02/npdc_annual_report_r02_2-02.pdf.]

Japan Meteorological Agency, 2016: Launch of provision of estimated weather distribution products. Technical Information on Distribution Materials. 422, Japan Meteorological Agency, 11 pp (in Japanese). [Available at https://www.data.jma.go.jp/add/suishin/jyouhou/pdf/422.pdf.]

Japan Meteorological Agency, 2022: Meso-scale numerical prediction system vertical layer augmentation, forecast range extension, and physical process improvements. Numerical Prediction Development Center Annual Report 2021. Japan Meteorological Agency, 92-99 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/npdc/r03/npdc_annual_report_r03_4-03.pdf.]

Japan Meteorological Agency, 2023a: NWP application products. Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency. Japan Meteorological Agency, 157-188. [Available at https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline-latest-nwp/pdf/outline2023_04.pdf.]

Japan Meteorological Agency, 2023b: Data Assimilation Systems. Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency. Japan Meteorological Agency, 14 pp [Available at https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline-latest-nwp/pdf/outline2023_02.pdf.]

Japan Meteorological Agency, 2023c: Numerical Weather Prediction Models. Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency. Japan Meteorological Agency, 53-156 [Available at https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline-latest-nwp/pdf/outline2023_0

3.pdf.]

Kawano, K., M. Ujiie, M. Kunii, and S. Nishimoto, 2019: Meso-scale ensemble prediction system. Textbook for Numerical Weather Prediction. 52, Japan Meteorological Agency, 1-15 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwptext/52/chapter1.pdf.]

Kingma, D. P., and J. L. Ba, 2015: Adam: A method for stochastic optimization. Conference paper at the Third International Conference on Learning Representations 2015, San Diego, U.S.A., 15 pp, doi:10.48550/arXiv.1412.6980.

Klein, W. H., and H. R. Glahn, 1974: Forecasting local weather by means of Model Output Statistics. Bull. Amer. Meteor. Soc., 55, 1217-1227, doi:10.1175/1520-0477(1974)055<1217:FLWBMO>2.0.CO;2.

Kudo 2022: Statistical Post-Processing for Gridded Temperature Prediction Using Encoder–Decoder-Based Deep Convolutional Neural Networks, J. Meteorol. Soc. Japan., 100, 2019-232, doi:10.2151/jmsj.2022-011.

Kuroki, Y., 2017: Improvement of gridded temperature guidance and changes of guidance for snowfall amount and categorized weather. Textbook for Numerical Weather Prediction. 50, Japan Meteorological Agency, 94-101 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwptext/50/chapter4.pdf.]

Kusabiraki, H., and S. Moriyasu, 2013: Verification in the operational numerical weather prediction models. Report of Numerical Prediction Division. 59, Japan Meteorological

Agency, 16-24 (in Japanese).

Kusabiraki, H., 2020: Radiation. Report of Numerical Prediction Division. 66, Japan Meteorological Agency, 61-68 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwpreport/66/chapter2.pdf.]

Météo-France, 2015: JOINT WMO TECHNICAL PROGRESS REPORT ON THE GLOBAL DATA PROCESSING AND FORECASTING SYSTEM AND NUMERICAL WEATHER PREDICTION RESEARCH ACTIVITIES FOR 2015, Météo-France. [Available at https://wmoomm.sharepoint.com/:w:/s/wmocpdb/EfAO2BxjA7NDgdRscNaL488Brn1XlsY87wefjuIq7uRBug.]

Météo-France, 2020: JOINT WMO TECHNICAL PROGRESS REPORT ON THE GLOBAL DATA PROCESSING AND FORECASTING SYSTEM AND NUMERICAL WEATHER PREDICTION RESEARCH ACTIVITIES FOR 2015, Météo-France. [Available at https://wmoomm.sharepoint.com/:b:/s/wmocpdb/EQOKo5TMgDRAjyv3SFawEqIBGl4OKHS4txHR_PufRbIWTA.]

Met Office, 2015: JOINT WMO TECHNICAL PROGRESS REPORT ON THE GLOBAL DATA PROCESSING AND FORECASTING SYSTEM AND NUMERICAL WEATHER PREDICTION RESEARCH ACTIVITIES FOR 2015, Met Office. [Available at https://wmoomm.sharepoint.com/:b:/s/wmocpdb/EfYms7k874hOry4vAkDu-MoBOVC9363_VDOrvxJM-Sar8g.]

Miura, H., and Y. Ohashi, 2017: Influences of the relocations of the Okayama Local

Meteorological Observatory on the measured air-temperature. Naturalistae, 21, 7-15.

Nair, V., and G. E. Hinton, 2010: Rectified linear units improve restricted Boltzmann machines. Proceedings of the Twenty-seventh International Conference on Machine Learning (ICML-10), Haifa, Israel, 807-814.

Sannohe, Y., 2018: Temperature guidance. Report of Numerical Prediction Division. 64, Japan Meteorological Agency, 132-143 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwpreport/64/chapter4.pdf.]

Segami, T. M. Obayashi, M. Kunitsugu, and T. Fujita, 1995: Kalman filter. Textbook for Numerical Weather Prediction, 28, Japan Meteorological Agency, 66-78 (in Japanese)

Sheridan, P., S. Smith, A. Brown, and S. Vosper, 2010: A simple height-based correction for temperature downscaling in complex terrain. Meteor. Appl., 17, 329-339, doi:10.1002/met.177.

Shi, X., Z. Chen, H. Wang, D. Yeung, W. Wong, W. Woo, 2015: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, NeurIPS 2015, 7-12 December 2015, Montreal, Quebec, Canada. Cortes, C., N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (eds.), 802-810, doi:10.48550/arXiv.1506.04214.

Suzuki, K., T. Iwasaki, and T. Yamazaki, 2021: Analysis of systematic error in Numerical Weather Prediction of coastal fronts in Japan's Kanto Plain. J. Meteor. Soc. Japan, 99, 27-47, doi:10.2151/jmsj.2021-002.

Takada, S., 2018a: Introduction to guidance. Report of Numerical Prediction Division. 64, Japan Meteorological Agency, 3-8 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwpreport/64/chapter1.pdf.]

Takada, S., 2018b: Support for NWP updates. Report of Numerical Prediction Division. 64, Japan Meteorological Agency, 88-90 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwpreport/64/chapter3.pdf.]

Takada, S., 2018c: The impact of and response to observatory relocation. Report of Numerical Prediction Division. 64, Japan Meteorological Agency, 91-93 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/nwpreport/64/chapter3.pdf.]

Veira, A., R. Hess, S. Trepte, G. Vogt, and B. Reichert, 2017: Model Output Statistics for Point Forecasts at Deutscher Wetterdienst: Current Status and Future Developments. European Conference for Applied Meteorology and Climatology 2017, European Meteorological Society, EMS2017-378. [Available at http://meetingorganizer.copernicus.org/EMS2017/EMS2017-378-1.pdf.]

Wakayama, I., T. Imai, T. Kitamura, and K. Kobayashi, 2020: About estimated weather distribution. Weather service bulletin, 87, Japan Meteorological Agency, 1-18 (in Japanese). [Available at https://www.jma.go.jp/jma/kishou/books/sokkou/87/vol87p001.pdf.]

Wilks, D., 2011: Statistical Methods in the Atmospheric Sciences. 3rd ed. International

Geophysics Series, Vol. 100, Academic Press, 704 pp.

Wilson, L. J. and M. Vall´ee, 2002: The Canadian Updateable Model Output Statistics (UMOS) System: Design and Development Tests. Wea. Forecasting, 17, 206–222, doi:10.1175/1520-0434(2002)017<0206:TCUMOS>2.0.CO;2.

Zurndorfer, E. A., J. R. Bocchieri, G. M. Carter, J. P. Dallavalle, D. B. Gilhousen, K. F. Hebenstreit, and D. J. Vercelli, 1979: Trends in comparative verification scores for guidance and local aviation/public weather forecasts. Mon. Wea. Rev., 107, 799-811, doi:10.1175/1520-0493(1979)107<0799:TICVSF>2.0.CO;2.

List of Figures

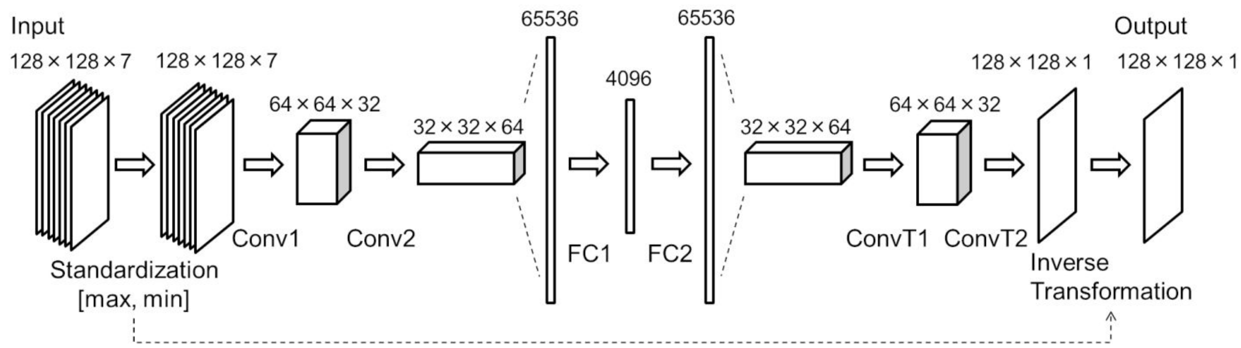

Fig. 1 Schematic diagram of the deep convolutional neural network proposed in Kudo (2022). Only the input/output image size differs from that of Kudo (2022). The details of the operation units, such as Conv1 and Conv2, are described in Table 1.

Fig. 2 The six target areas (jp01-06) covering the major regions of Japan. This map is based on *the Digital Map 5000000 Japan and Its Surroundings (Integration)* published by the Geospatial Information Authority of Japan. The bathymetric contours are derived from *the General Bathymetric Chart of the Oceans (GEBCO) Digital Atlas* published by the British Oceanographic Data Centre (BODC) for the Intergovernmental Oceanographic Commission (IOC) and the International Hydrographic Organization (IHO). The shoreline data are derived from *the Vector Map Level 0 (VMAP0)* of the National Imagery and Mapping Agency of the United States and the United States Geological Survey (USGS)

Information Services.

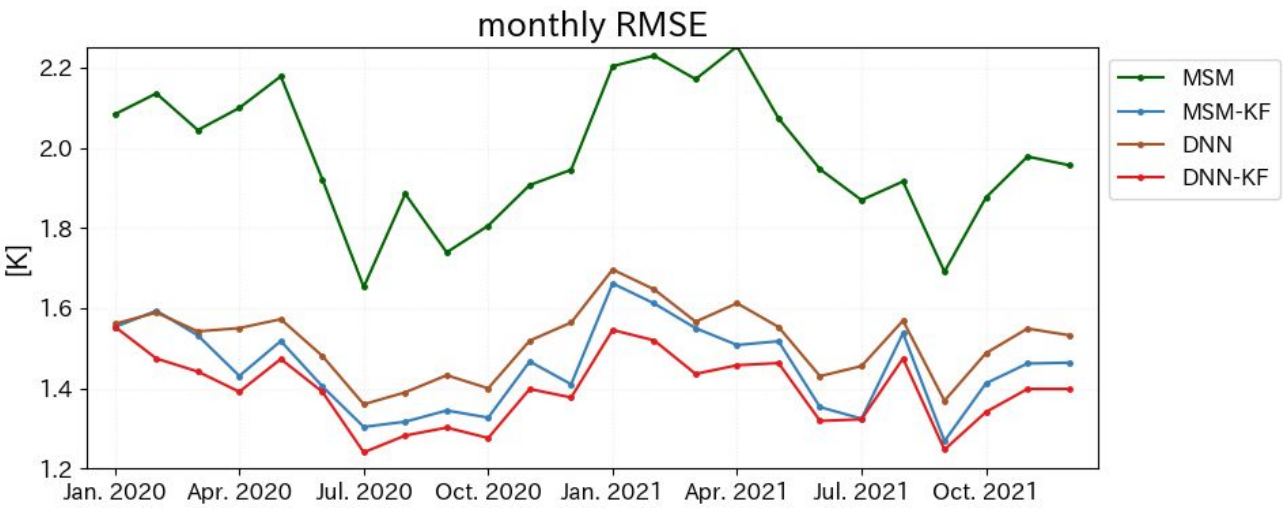

Fig. 3 Monthly averaged RMSEs at each observatory of temperature forecasts for the interpolated MSM, operational point-like guidance (point-like MSM-KF), interpolated DNN-based gridded prediction (DNN), and DNN-based point-like guidance forecast (DNN-KF).

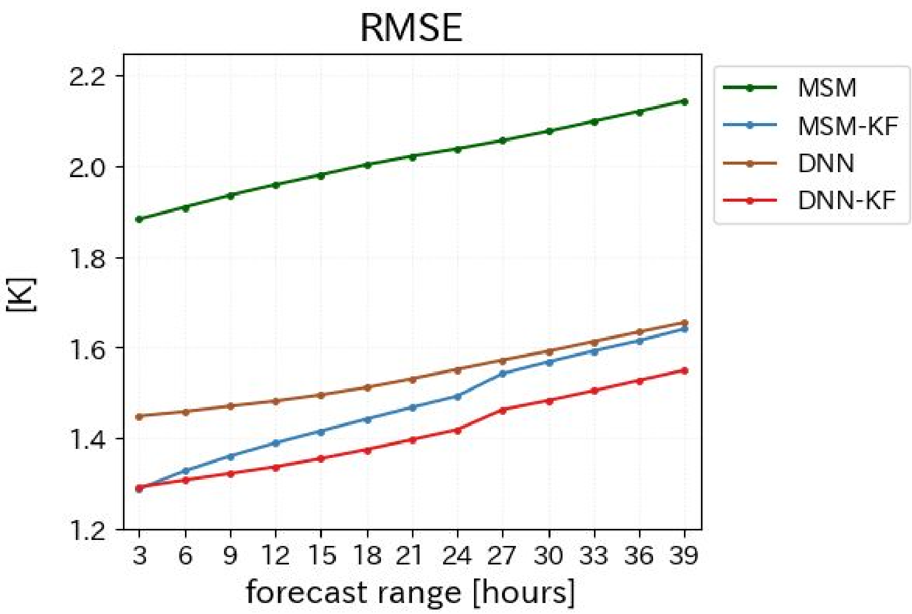

Fig. 4 Average RMSEs classified by forecast lead times for the interpolated MSM, point-like

MSM-KF, interpolated DNN, and DNN-KF from January 1 to December 31, 2021.

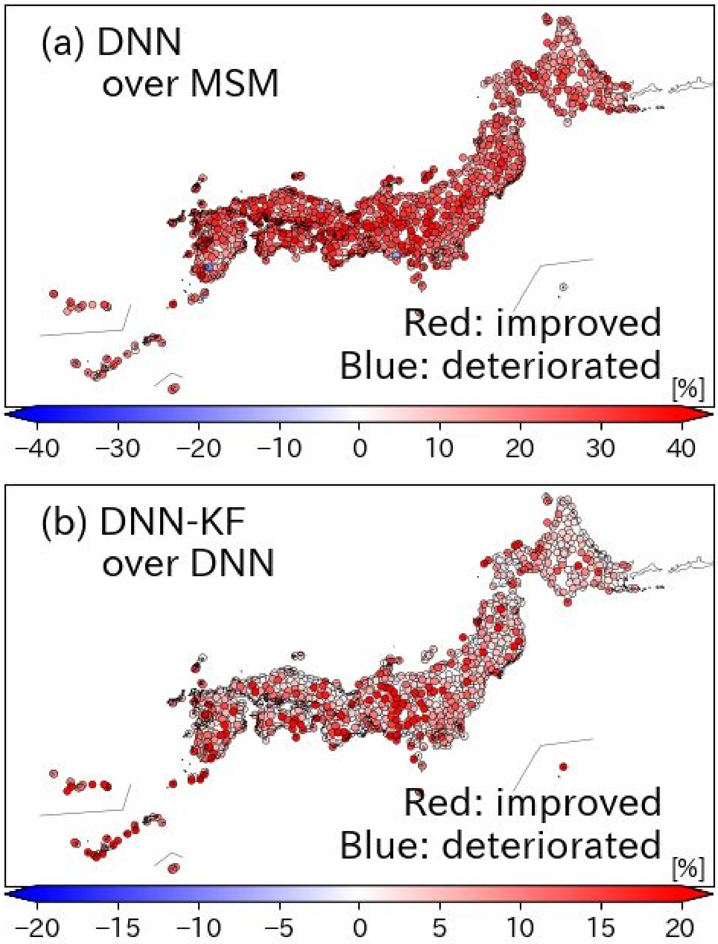

Fig. 5 The relative improvements in (a) the interpolated DNN over the interpolated MSM and (b) the DNN-KF over the interpolated DNN at each observatory. Red (blue) circles represent improved (deteriorated) observatories. The test period is from January 1 to December 31, 2021.

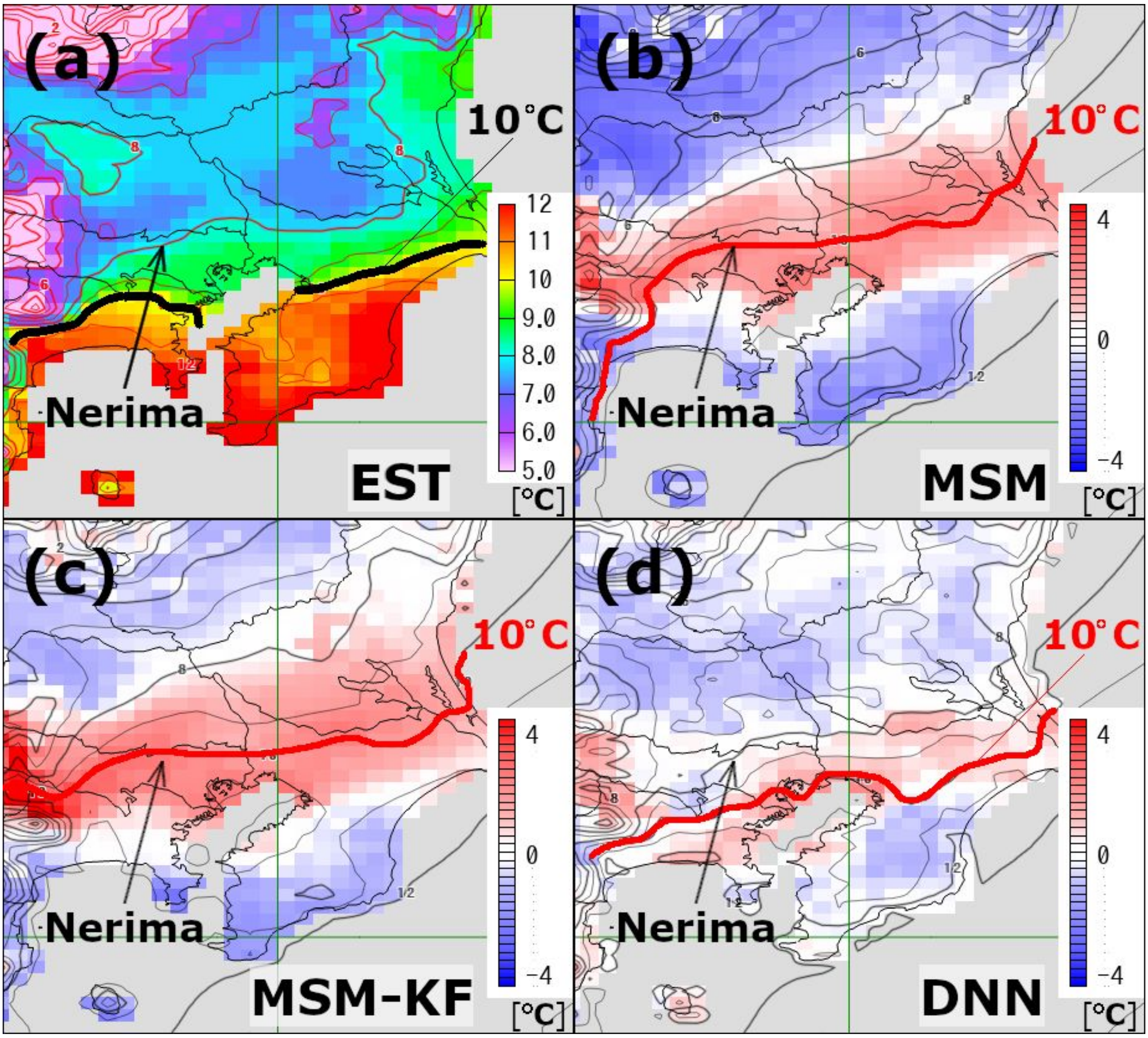

Fig. 6 (a) Surface temperatures in the Kanto (jp03) region at 15 LST on December 29, 2021 for the real-time estimated surface temperature (EST) distribution provided by the JMA (contours and color shading), (b) the temperature forecast of the MSM (contours) and its differences from the EST (color shading), (c) the forecast of the gridded MSM-KF (contours) and its differences from the EST (color shading), and (d) the forecast of the DNN (contours) and its differences from the EST (color shading). The forecasts are initialized at 21 LST on December 28, 2021.

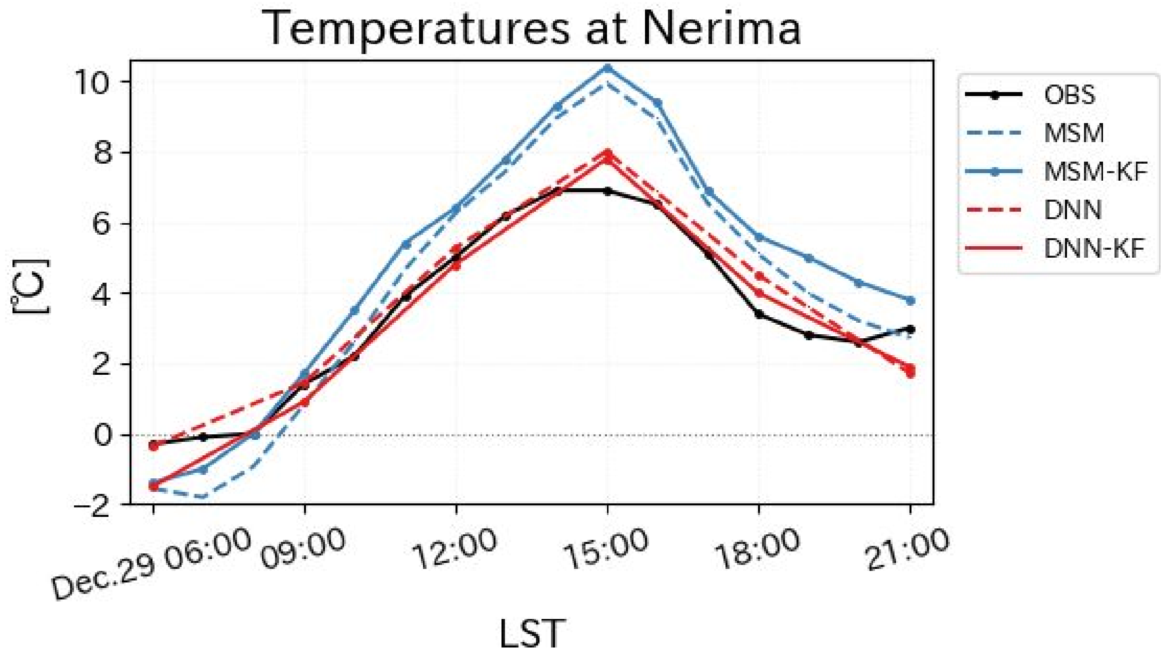

Fig. 7 Time series of temperatures for in-situ observations (OBS), the interpolated MSM forecast, point-like MSM-KF, interpolated DNN, and DNN-KF at Nerima (shown in Fig. 6), initiated at 21 LST on December 28, 2021.

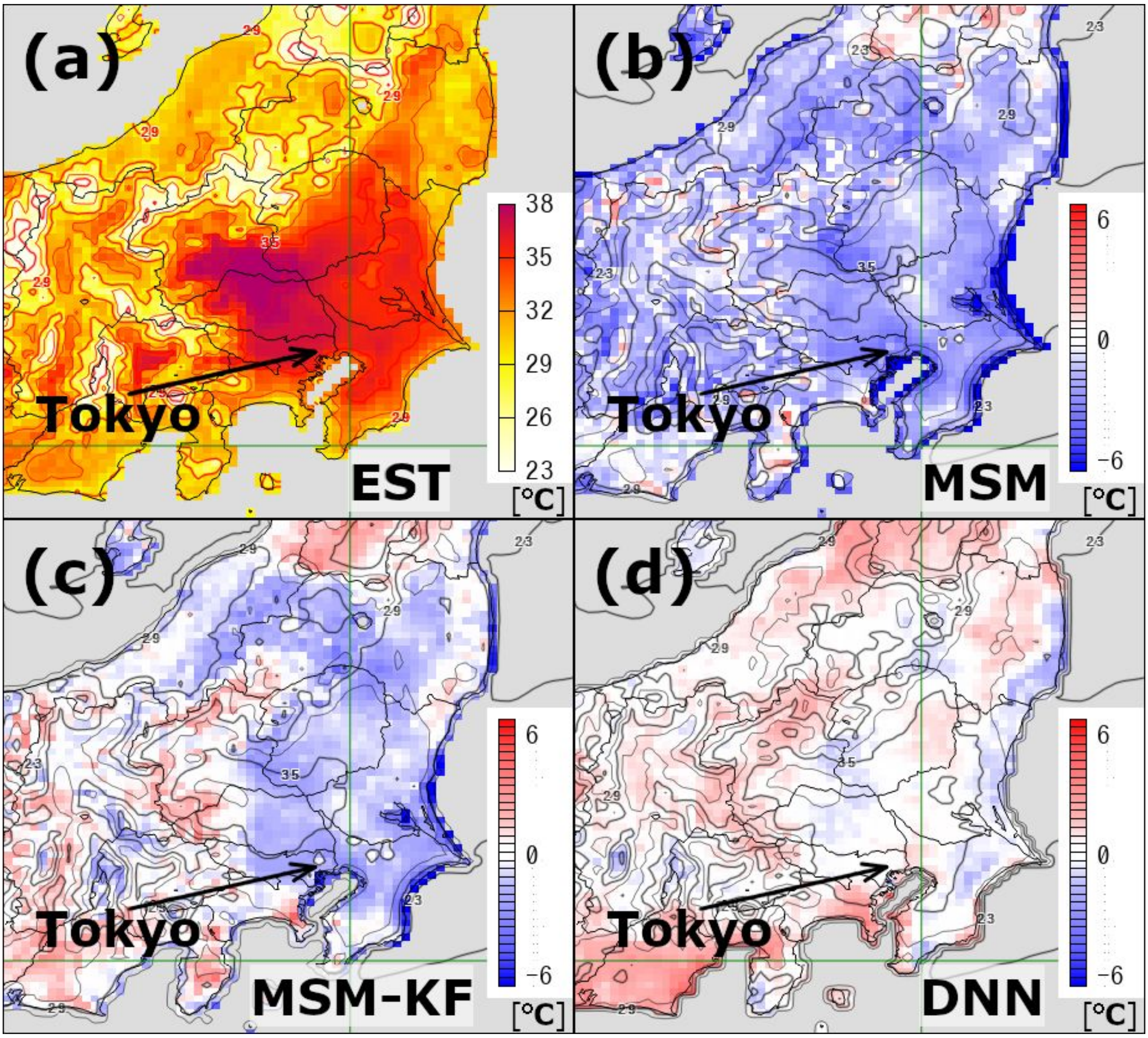

Fig. 8 Same as Fig. 6 but for the projection time at 12 LST on July 1, 2022 and the initial time at 15 LST on June 30, 2022.

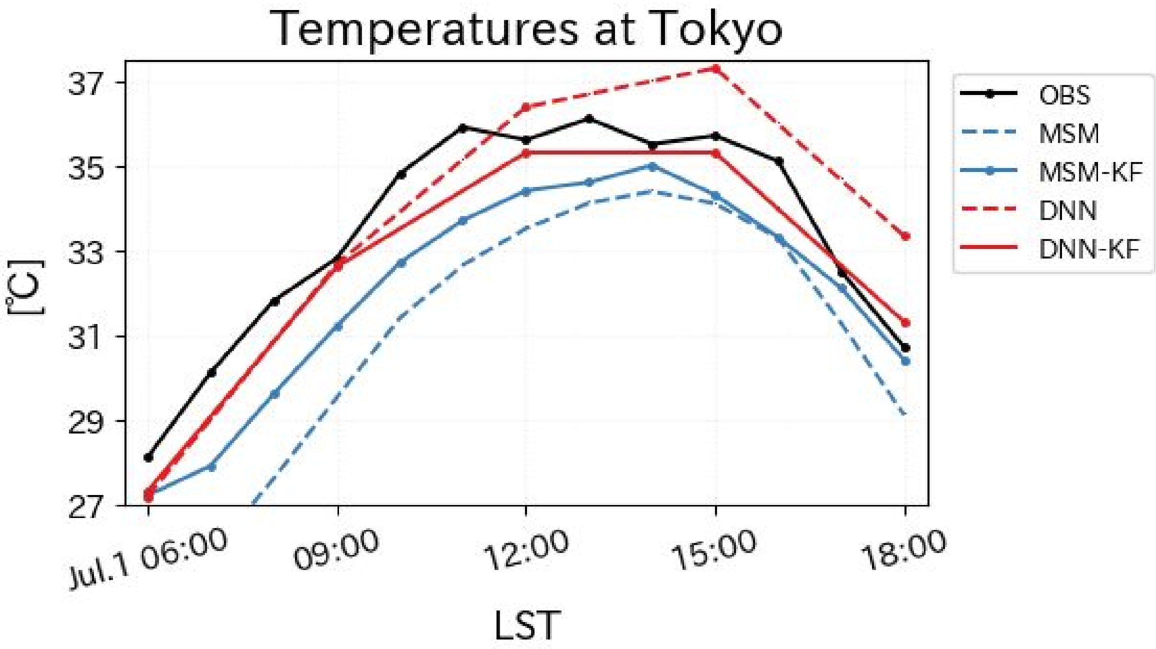

Fig. 9 Same as Fig. 7 but for the initial time at 15 LST on June 30, 2022 at Tokyo (shown in Fig. 8).

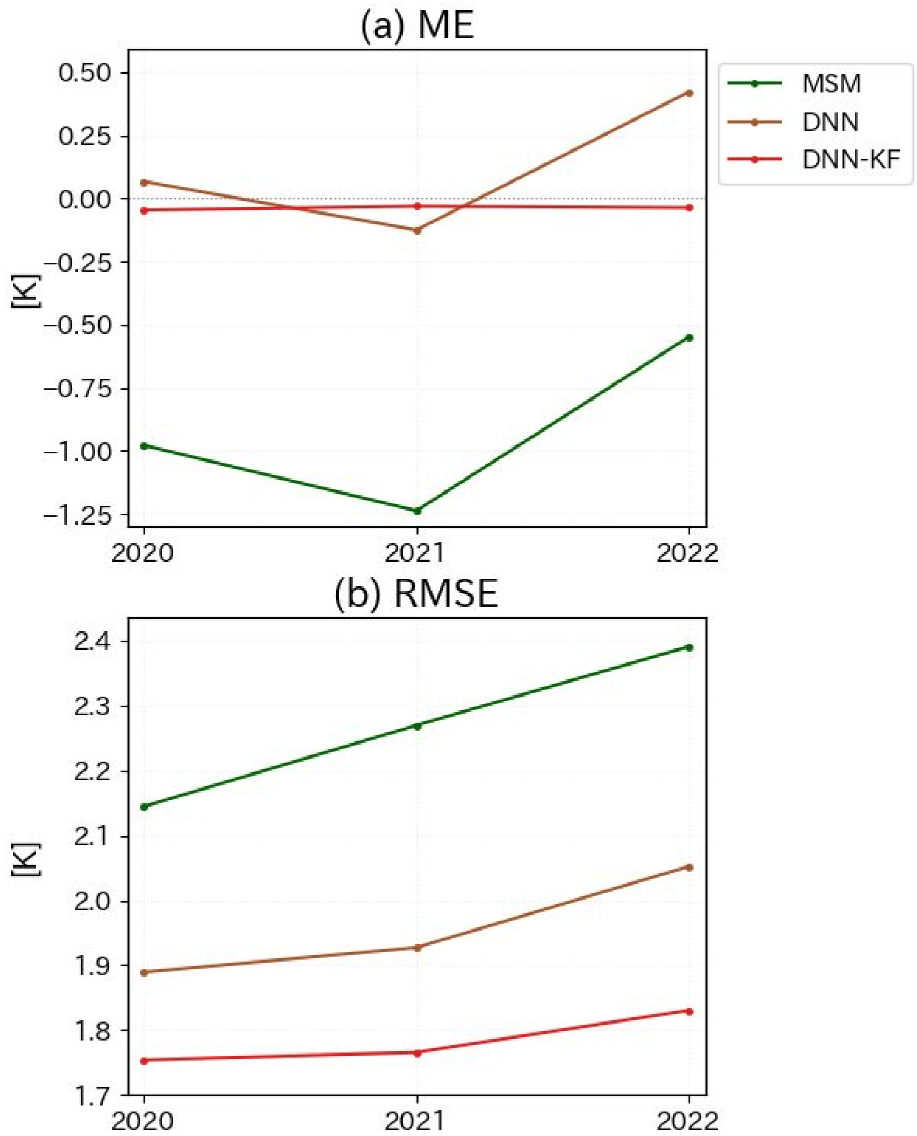

Fig. 10 Interannual changes in (a) MEs and (b) RMSEs of temperature forecasts at each observatory in the Kanto region for the interpolated MSM, interpolated DNN, and DNN-KF predictions at 15 LST from 2020 to 2022 in summer.

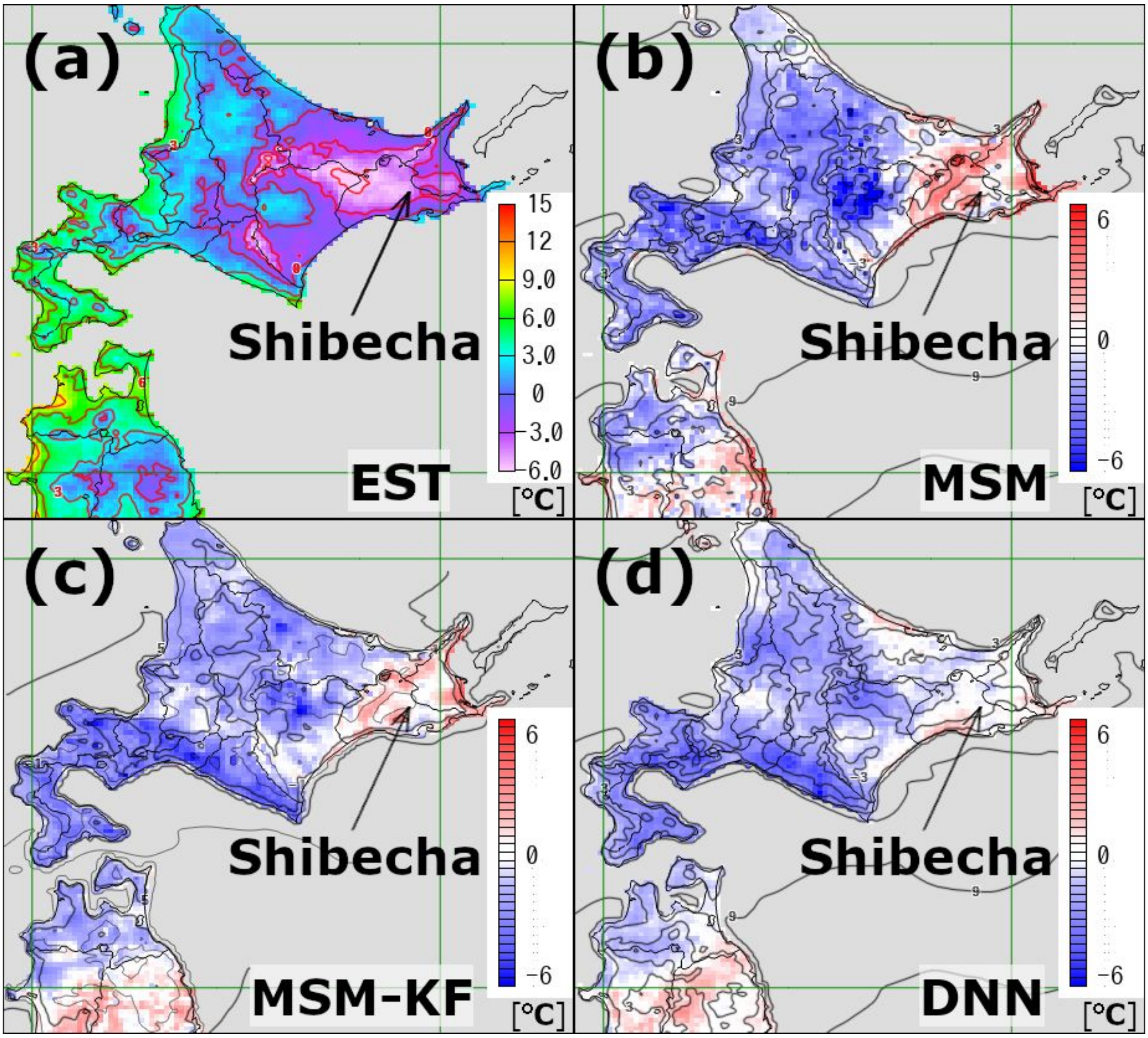

Fig. 11 Same as Fig. 6 but for the northernmost region of Japan (jp01, jp02) with a projection time at 06 LST on November 16, 2021 and an initial time at 21 LST on November 14, 2021.

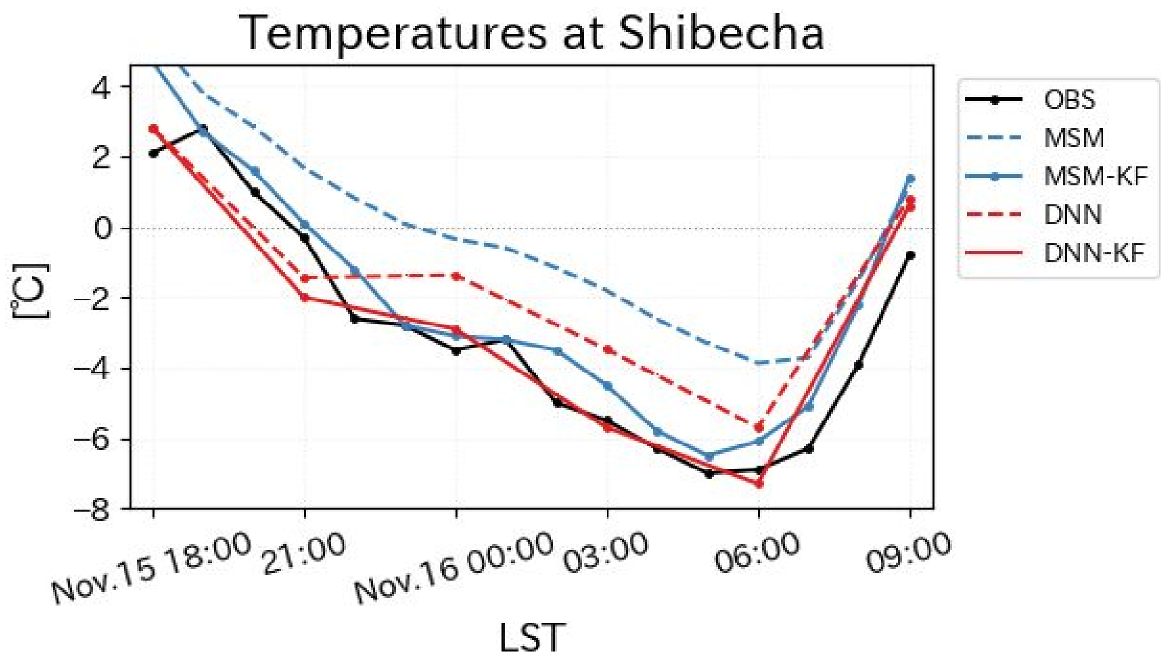

Fig. 12 Same as Fig. 7 but for the initial time at 21 LST on November 14, 2021 at Shibecha (shown in Fig. 11).

List of Tables

Table 1. Functions and parameters used in the network shown in Fig. 1.

Unit	Function	Parameters
Conv1	Conv2d	kernel_size = 5, stride = 1, padding = 2, number of channels: 7 → 32
	MaxPool2d	kernel_size = 2, stride = 2
	BatchNorm2d	number of channels: 32
	ReLU	
Conv2	Conv2d	kernel_size = 5, stride = 1, padding = 2, number of channels: 32 → 64
	MaxPool2d	kernel_size = 2, stride = 2
	BatchNorm2d	number of channels: 64
	ReLU	
FC1	Linear	number of units: 65536 → 4096
	BatchNorm1d	number of units: 4096
	ReLU	
FC2	Linear	number of units: 4096 → 65536
	BatchNorm1d	number of units: 65536
	ReLU	
ConvT1	ConvTranspose2d	kernel_size = 2, stride = 2, padding = 0, number of channels: 64 → 32
	BatchNorm2d	number of channels: 32
	ReLU	
ConvT2	ConvTranspose2d	kernel_size = 2, stride = 2, padding = 0, number of channels: 32 → 1
	BatchNorm2d	number of channels: 1
	Sigmoid	

Table 2. Time periods for training, validation, fine-tuning, and testing.

Dataset period	DNN-based gridded prediction (DNN)	DNN-based point-like guidance forecast (DNN-KF)
Oct. 8 in 2010 – Dec. 31 in 2018	training	-
Jan. 1 – Dec. 31 in 2019	validation, fine-tuning	-
Jan. 1 – Dec. 31 in 2020	test	training
Jan. 1 – Dec. 31 in 2021	test	test