Bias correction for spatially interpolated daily mean air temperature during winter in eastern Hokkaido using multimodal machine learning

Keach Murakami\textsuperscript{a,}\textsuperscript{†}, Tomoyoshi Hirota\textsuperscript{ab}, Seiji Shimoda\textsuperscript{c} and Tomotsugu Yazaki\textsuperscript{d}

\textsuperscript{a}Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, 1 Hitsujiyagai, Toyohira, Sapporo, Hokkaido 062–8555, Japan
\textsuperscript{b}Present address: Faculty of Agriculture, Kyushu University, 774 Motooka, Nishi, Fukuoka, Fukuoka 819–0395, Japan
\textsuperscript{c}Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, 9–4 Shinsei-Minami, Memuro, Kasai, Hokkaido 082–0081, Japan
\textsuperscript{d}School of Agriculture, Meiji University, 1–1–1 Higashi-Mita, Tama, Kawasaki, Kanagawa 214–8571, Japan

Abstract

Interactions between boundary layer wind and topography form non-uniform air temperature distributions in cold and snow-covered regions. Because of this heterogeneity, spatially interpolated air temperatures sometimes deviate from observed values. To evaluate the quality of spatially interpolated daily mean temperatures ($T_{\text{int}}$) provided by a 1 km gridded meteorological data service (Ohno \textit{et al.}, 2016), we collected observed temperatures ($T_{\text{obs}}$) obtained at meteorological observation sites located near farmland in the Tokachi and Okhotsk regions—in eastern Hokkaido, Japan—in winter (October–March) and revisited the bias in the interpolated temperatures ($dT$). The root-mean-square error (RMSE) of $T_{\text{int}}$ obtained at 88 sites was 1.16°C, and the absolute median $dT$ values were greater than 1°C at 14 sites. The variance of $dT$ was greater on cold and calm days, suggesting the involvement of radiative cooling and the accumulation of cold air parcels. To correct $T_{\text{int}}$ by estimating $dT$ at a given site by considering the formation mechanisms of the temperature distributions, we attempted to develop a multimodal machine learning model that had four predictors: surface and boundary layer meteorological data and topographical and geographical features around each site. To analyze the influence of the spatial extent of the topography and geography around each site, we compared models having these predictors with various sizes of the region of interest (ROI). By training the models and applying them to an independent test dataset, it has been shown that bias correction using models with a small topographical ROI ($\times 30$ km or smaller) reduced the RMSE. The RMSE of the test dataset decreased by $\approx 0.1$°C via the application of a nested model, suggesting the potential usefulness of the presented approach for locally confined meteorological events. However, the biases were increased at several sites by application of the models, thus implying that further improvement is essential for practical use.

\textbf{Key words:} Boundary layer wind, Downscaling, Geography, Meso-alpha scale meteorology, Topography

1. Introduction

The accurate and fine spatial distributions of meteorological elements facilitate IT-based agricultural practices (e.g., Yazaki \textit{et al.}, 2013; Yazaki and Hirota, 2019). Since the number of meteorological stations and their spatial coverage are limited because of labor and financial costs, attempts have been conducted to obtain finely resolved maps of meteorological elements (e.g., Pielke and Wilby, 2012). Statistical downscaling is a widely adopted cost-effective method used to enhance the spatial resolution of meteorological elements. In Japan, for example, daily meteorological elements are estimated at a grid spacing of $\approx 1$ km by using statistical downscaling, which are delivered by the Agro-Meteorological Grid Square Data System (AMGSD; Ohno \textit{et al.}, 2016; Ohno and Sasaki, 2019).

Statistical downscaling uses transfer functions (e.g., regression relationships) representing observed relationships between larger-scale atmospheric variables and local quantities, such as daily precipitation and/or temperatures (Pielke and Wilby, 2012). Besides large-scale atmospheric variables, local topographical features (e.g., relief and gradient) have also been used as predictor variables (e.g., Pardo-Iguzquiza, 1998; Ninyerola \textit{et al.}, 2000; Sameshima \textit{et al.}, 2008; Japan Meteorological Agency [JMA], 2011; Ueyama, 2013; Feidas \textit{et al.}, 2014). The downscaling procedure deduces a generic relationship from data obtained at several sites, leading to regression toward the mean. In other words, the possibility that extreme values may appear in the interpolated area is always smaller than in reality.

In terms of air temperatures in Japan, extremely low temperatures have sometimes been observed in northern regions during winter having thick and dry snowpack if the topographic and boundary layer conditions are suitable for strong cooling (Kondo, 2000). The winter air temperature affects the survival ratio and productivity of overwintering wheat (e.g., Fowler \textit{et al.}, 1999; Shimoda \textit{et al.}, 2015) and fruit trees (e.g., Ashworth and Wisniewski, 1991). For example, the production of wine grapes has become widespread in Hokkaido during these decades (Hirota \textit{et al.}, 2017), and the winter air temperature determines suitable areas for wine grape cultivation to avoid frost damage (Nemoto \textit{et al.}, 2016). The winter air temperature also affects the depth of soil frost and the emergence of volunteer potatoes,
which reduces light interception of the succeeding crops and harbor insects, diseases, and nematodes (e.g., Hirota et al., 2011; Hirota and Kobayashi, 2019). Therefore, the winter air temperature is an influential meteorological factor for agriculture in Hokkaido. Yazaki et al. (2017) reported that the winter air temperature in the Tokachi Plain, which is situated in the eastern part of Hokkaido Island in northern Japan, was not spatially uniform. They measured air temperatures at several sites and found that the difference in the daily minimum values at two close sites ~5 km apart was sometimes greater than 10°C. By applying a nonhydrostatic model to the region and modifying the boundary layer wind conditions, further analysis suggested that the wind and topography caused the air temperature to form a heterogeneous spatial distribution (Fukushima et al., 2019).

Although several earlier studies that considered downscaling quantified the effect of local topographical features on the daily mean air temperature by applying stepwise multiple regression models (e.g., Kanno, 1997; Sameshima et al., 2008; Ueyama, 2008; JMA, 2011), interactions among the boundary layer wind, topography, and air temperature were not incorporated. The multiple regression conducted to analyze a monthly bias suggested that the influential topographical features differed among the months (JMA, 2011), presumably depending on the seasonal wind condition. As suggested by these results, it is necessary to develop a method that enables the accurate interpolation of air temperatures in cold and snow-covered regions by untangling the complex interactions between the boundary layer wind and topography around a given site.

To address this issue, in this work, we adopted a machine learning method. Several recent studies have investigated the application of machine learning techniques to weather forecasting (e.g., Shi et al., 2015; Yonekura et al., 2018; Liu et al., 2019). Another important moiety to be addressed is spatial estimation, namely, downscaling. Several studies have attempted to apply these techniques to meteorological downscaling and demonstrated their effectiveness in terms of air temperatures (Baño-Medina et al., 2019; Sachindra and Kanae, 2019), precipitation (e.g., Wilby et al., 1998; Vandal et al., 2017; Misra et al., 2018; Baño-Medina et al., 2019; Pan et al., 2019), and wind speeds (Li, 2019). Most of these earlier studies focused on obtaining variables at a grid spacing of approximately 10–50 km by downscaling large-scale variables simulated by general circulation models. Therefore, we should test the applicability of machine learning methods to the estimation of air temperatures that are heterogeneous within a small scale (e.g., a grid spacing of ~1 km). Furthermore, the earlier studies used machine learning to obtain finely resolved data that are similar to super-resolution images (e.g., Dong et al., 2014). The applicability of machine learning methods to the bias correction for values obtained by downscaling has not been examined.

The present study aimed (1) to revisit the biases in spatially interpolated air temperatures in Hokkaido during winter, where the spatially heterogeneous temperature distributions were formed and (2) to propose a method that estimates the deviations of interpolated daily mean air temperatures at a given site by considering the interactions between the boundary layer wind and topography. To incorporate these complex interactions, we attempted to develop a multimodal machine learning model using meteorological and topographical predictors.

2. Materials and Methods

2.1 Study domain and meteorological observation

We focused on two Japanese regions located in eastern Hokkaido: the Tokachi and Okhotsk regions (Fig. 1). These regions are predominated by farmlands consisting of upland crops, such as wheat, potato, and several other crops. In the present study, we collected daily mean air temperatures measured at meteorological observation sites operated during winter (from October to March in 2016/2017–2018/2019). These sites were maintained by an incorporated foundation (Japan Weather Association, Tokyo, Japan; JWA) and an agricultural consulting firm (Agriweather Inc., Sapporo, Japan; AGW), which were funded by the respective

Fig. 1. Maps showing the study domain and meteorological observation sites maintained by AGW (◇), JWA (▽), and HARC (○). The color represents the site elevation [m]. [A color version of the figure is available online]
municipalities and farmers’ cooperatives. Since these sites were mainly operated for agricultural usage, they were surrounded by farmlands. At the JWA sites \((N = 50)\), the air temperatures were measured at each meteorological station using an installed forced ventilated thermometer in a ventilation tube at 1 h time intervals. At the AGW sites \((N = 18)\), the air temperatures were measured with a temporarily installed weather station \((\text{WeatherBucket}, \text{SEC Co. Ltd., Hakodate, Japan})\) at 10 min time intervals. Besides these sites, we also measured the air temperatures in local farmers’ fields \((\text{HARC}, \text{\text{N} = 20}}, \text{2010/2011–2018/2019})\). The air temperatures were measured using a PT-100 sensor \((\text{HMP155; Vaisala Corp., Helsinki, Finland})\) in a ventilation tube at a height of 1.9 m at 1 min time intervals as described by Yazaki et al., 2017 or using portable thermometers \((\text{TR-52, T&D Corporation, Tokyo, Japan})\) in a natural ventilated shelter at a height of 1.8 m at 1 h time intervals \((N = 6 \text{ and } 14, \text{respectively})\). We preliminarily compared daily mean air temperatures obtained by natural and forced ventilation systems at four sites for 2 years and confirmed that the root-mean-square error (RMSE) caused by the use of a natural ventilation system was \(-0.36^\circ \text{C}\). Sameshima et al., 2007 also reported that the RMSE in daily minimum air temperature was small \((0.3^\circ\text{C})\). We conducted quality control on the basis of the daily mean, minimum, and maximum air temperatures, and removed anomalies \((-6\% \text{ of total records})\).

### 2.2 Datasets

In the present study, the measured daily mean air temperatures \((T_{\text{obs}})\) were compared with the interpolated estimates provided by AMGSD \((T_{\text{am}})\) in a \(1 \times 1 \text{ km grid}\), and the calculated differences in the temperatures \((dT = T_{\text{obs}} - T_{\text{am}})\) were used as the predictands for machine learning. The following datasets were used for machine learning.

A surface meteorological dataset was generated using AMGSD. Briefly, this system calculates a \(1 \times 1 \text{ km grid}\) of meteorological elements on the basis of climatological normal values, deviations from the normal values, and elevation corrections for several elements including \(T_{\text{am}}\). The climatological normal values were estimated at a grid spacing of 1 km by the JMA based on the multiple stepwise regression method with topographical and geographical predictors (JMA, 2011). As described by Ohno et al., 2016, the deviations from the climatological normal values were estimated by interpolating the deviations observed at the automated meteorological data acquisition system \((\text{AMeDAS})\) stations operated by JMA. Twelve meteorological elements at the corresponding grid to each site, as well as its elevation, were used as predictors for the model. The site elevations were obtained via API (Geospatial Information Authority of Japan, 2013; http://maps.gsi.go.jp/development/elevation_s.html).

A boundary layer meteorological dataset was generated from the global spectral model developed by JMA. To incorporate the stability of the boundary layer atmosphere into the model, the horizontal and vertical wind speeds and air temperature distributions above the areas surrounding Japan \((100–170^\circ \text{E in longitude and } 10–70^\circ \text{N in latitude})\) were used as predictors. We used the values at 700, 850, and 925 hPa heights at JSTD. The data were downloaded from a Web server maintained by a scientific community (GFD-DENNOU Club; https://www.gfd-dennou.org/index.html.en).

Topographical and geographical datasets were generated from a global digital surface model \((\text{ALOS World 3D–30 m; Japan Aerospace Exploration Agency, 2019})\) and a land use classification dataset \((\text{Land Use Fragmented Mesh; National Land Numerical Information, 2016})\). The grid spacings for these datasets were 30 m and 1 km, respectively. The fractional areas occupied by bodies of water \((\text{sum of the rivers, lakes, and seawater})\) and urban land \((\text{sum of land used for building, trunk transportation land, and other lands})\) were calculated from the land use classification dataset and used as predictors. To incorporate the interactions with the wind direction, these two-dimensional matrices were rotated according to the most frequent daily wind direction at the closest AMeDAS station. In other words, the wind direction was fixed in terms of rows and columns of the matrices. When the closest AMeDAS station was calm throughout a day, the corresponding record \((i.e., \text{the surface and boundary layer meteorological data and the geographical and topographical features on the day})\) were removed \((-0.01\% \text{ of the total records})\).

Notably, the air temperature data observed at the AMeDAS stations were not used for the model because AMGSD estimated the temperature by interpolating deviations observed at AMeDAS stations. The \(dT\) values at the AMeDAS stations were always \(-0\) because of this procedure, and thus, the data were not appropriate for machine learning.

### 2.3 Architecture of machine learning model and procedures for training and evaluation

In the present study, we attempted to develop a machine learning model that estimates \(dT\) at a given site using multimodal predictors \((\text{Fig. 2})\). The software R (ver. 3.6.2; R Core Team, 2019) was used for data preprocessing and as a frontend for the machine learning using a neural network library Keras (ver. 2.2.5.0; Allaire and Chollet, 2018); for the backend, TensorFlow (ver. 2.0.0; Allaire and Tang, 2018) was used. A single GPU \((\text{TITAN RTX, NVIDIA Corp., Santa Clara, CA})\) with a 24GB memory was used.

We applied the principal component analysis (PCA) and convolutional neural network (CNN) to extract features from multimodal datasets \((\text{Fig. 2})\). Features from the vector predictor \((i.e., \text{AMGSD dataset})\) were extracted by PCA using psych::principal function \((\text{Revelle, 2018})\) via the software R to avoid multicollinearity. Four principal components with eigenvalues greater than unity were selected. Features from the two-dimensional predictors \((i.e., \text{topographical and geographical data and boundary layer meteorological data})\) were extracted using CNN layers and then converted into a vector format. The features extracted from all datasets were concatenated and subsequently fed to a downstream neural network to obtain an output. During the model training, parameters were determined to minimize the difference between this output and \(dT\). The model was trained using \(T_{\text{obs}}\) measured at the JWA and AGW sites and the corresponding records on the same day without using records on the days before. Out of 10,000 randomly sampled records, 9,000 and 1,000 records were used for training and validation, respectively. The models were trained using the stochastic gradient descent optimizer with a learning rate and momentum...
of 0.01–0.02 and 0.9, respectively, a log-cosh loss function, the mean absolute error as a metric function, a batch size of 32, and an epoch size of 120. The training was stopped when the validation losses had stopped decreasing for 20–50 epochs. The trained model was evaluated using an independent test dataset obtained at the HARC sites (2,942 records). The R code used to build, train, and test the present model is available (Doc. S1).

We conducted three experiments in the present study. In the first experiment, we examined an ordinary neural network model—three hidden layers with 64, 32, and 16 nodes—that took only the surface meteorological data as a predictor. This experiment was conducted to confirm that the surface meteorological data were not enough to explain the bias. The model exhibited little responses and failed to correct the bias as expected data not shown.

In the second and third experiments, the boundary layer meteorological dataset (i.e., vertical and horizontal wind and air temperatures at 700, 850, and 925 hPa heights) and topographical dataset were fed as predictors of the machine learning model as well as the surface meteorological dataset. We also added a geographical dataset (i.e., water body coverage and urban land coverage) as a predictor to analyze the possible involvement of the land use around the sites. The size of the surrounding area (ROI) influencing the air temperature should vary among the predictors. In the second experiment, we investigated the influences of the ROIs of topography, water body coverage, and urban land coverage on the model performance.

We developed models with various combinations of ROIs and compared them as follows. Among the three two-dimensional predictors (i.e., topography, water body coverage, and urban land coverage), we set one predictor as a variable—ROI predictor, and its ROI was either ~6 × 6, 12 × 12, 30 × 30, 60 × 60, or 120 × 120 km. The ROIs of the other two predictors were fixed at either 30 × 30 or 120 × 120 km. The matrices were cropped from the datasets and resized into 128 × 128 px using nearest neighbor interpolation. We subsequently trained 30 = 3 predictors × 5 variable ROIs × 2 fixed ROIs models and evaluated RMSE values by using an independent test dataset. This approach was similar to the multiple stepwise regression method (e.g., JMA, 2011) to select appropriate ROIs for inputs (e.g., relief and gradient). In the third experiment, we attempted to develop an improved model to outperform the model developed in the second experiment. The model architecture (e.g., number of CNN layers, pixel resolution of the input predictors, and connection between layers) were exploratorily changed and trained. The trained models were evaluated using an independent test dataset, and the obtained RMSE values were compared to select an improved model.

3. Results and Discussion

3.1 Differences between the observed and interpolated daily mean air temperatures

To revisit the quality of the interpolated daily mean air temperature provided by AMGSD (Tint), we compared them
with observed values \((T_{\text{obs}})\) and evaluated the deviations \((dT = T_{\text{obs}} - T_{\text{int}})\). The RMSEs were 1.16°C in the present dataset. Among 88 sites, the median \(dT\) values were greater than +1°C in seven sites and smaller than −1°C in seven sites (Fig. 3). These systematic biases can result in significant errors in biological and physical process models, which provide accumulated daily outputs. For example, soil frost depth under thin snow, which has agricultural significances (see Introduction), has been modeled and practically controlled using daily mean air temperature as an input (Hirota et al., 2011).

The RMSE values were greater during mid-winter (Fig. 4). In earlier studies, the accuracies of the interpolated air temperatures were examined by leave-one-out cross-validation using values observed at AMeDAS stations (Seino, 1993; Ohno et al., 2016). In these studies, the temperature at a certain AMeDAS station was estimated by interpolating values observed at the other AMeDAS stations to calculate RMSE values. Seino (1993) applied this method to the Tokachi region and reported that the RMSE value of the daily mean air temperatures was approximately 0.7°C–1.2°C from October to March and was highest around January. Although the observed trend was consistent with this report, the RMSE values tended to be greater in the present study. Ohno et al. (2016) also reported that the RMSE values during winter (December–February) were greater than 1°C at several sites in Hokkaido. They further found that the RMSE values of the daily minimum values during winter was greater than 1°C at most sites in Hokkaido and mentioned the possible involvement of accumulated cold air parcels generated by katabatic drainage flow caused by local topological heterogeneity. The regions where the drainage flow reaches earlier night had lower daily minimum air temperatures (Fukushima et al., 2019). Since katabatic drainage flow was predicted to exist in the Tokachi and Okhotsk regions (Kimura, 1986) and was observed at several sites in the Tokachi Plain (Kimura, 1986; Fukushima et al., 2019), it could cause the greater RMSE values in the present dataset. Snowpack might also be related to the formation of heterogeneous air temperature distributions. The small thermal conductivity of snow (Kondo, 2000) could limit heat supply from the soil to the snow’s surface, thereby inducing a drastic reduction in the air temperature by radiative cooling during calm nights and accelerating the formation of patchy low-temperature zones above the snow’s surface. The RMSE values calculated by the interpolation with AMeDAS data were smaller in summer (Seino, 1993; Ohno et al., 2016) and in regions in Japan with thin or no snowpack during winter (Ohno et al., 2016), which supports the assertion that air temperature heterogeneity tends to form in cold and snow-covered regions. In a clear night, near-surface wind might also affect the temperature distribution (Yazaki et al., 2017). In regions where the near-surface wind is strong, sensible heat could be supplied to the surface, compensating for longwave radiative losses. Since the wind condition should be influenced by the geography and topography around the site (Fukushima et al., 2019), the formation of the patchy low-temperature zones would be either mitigated or emphasized depending on these local features.

To elucidate the causes of the deviations of \(T_{\text{int}}\), we analyzed the effects of the meteorological and topographical factors on the \(dT\) values. It was found that there was no clear relationship between \(dT\) and the distance to the closest AMeDAS station (Fig. 5). Considering that inaccurate estimates may originate during the interpolating process, any biases should increase along with the distance. The inaccurate \(T_{\text{int}}\) values should not be caused by the interpolation but rather by the heterogeneous temperature distributions in these regions. Seino (1993) mentioned that the estimated errors obtained by interpolating AMeDAS data could be overestimated because the distances between AMeDAS stations were longer than those for the interpolated areas located in the middle of AMeDAS stations. However, our result suggests that this discussion does not hold, at least in the domain of the present study.

The \(dT\) values tended to be smaller at sites with higher elevations (Fig. 6). The negative \(dT\) values at high elevation may suggest that the lapse rate for air temperatures adopted in AMGSD was underestimated. The lapse rate was fixed at 6.0 K km\(^{-1}\) in AMGSD, whereas a monthly regression analysis conducted in the north Japan region reported that the rate was 5.0–5.6 K km\(^{-1}\) during mid-winter.
(JMA, 2011). However, adjusting the lapse rate would not completely remove site-dependent systematic biases. There was no statistical correlation between the elevation and daily minimum air temperature (Yazaki et al., 2017), suggesting that the site-dependent systematic biases had a greater effect than the elevation. These biases should be corrected to obtain accurate \( T_{int} \) values.

### 3.2 Variability of the bias in the interpolated daily mean air temperatures

The variance of \( dT \) was greater on days with low daily mean air temperatures and low wind speeds than on days with high temperatures and high wind speeds (Fig. 7). In clear nighttimes on those days, cold air parcels could accumulate in basin areas because of the katabatic drainage flow, which may last for a long time. Although air parcels that formed in the early morning might disperse gradually, a strong positive correlation between the daily mean and minimum air temperatures (Yazaki et al., 2017) suggested that the minimum temperature was a major determinant of the mean temperature of that day. The accumulation of cold air parcels might cause both positive and negative \( dT \) variances via the following mechanisms. If a site is apt to fall into a basin, the \( T_{obs} \) values tend to be lower than \( T_{int} \) (i.e., negative \( dT \)). Conversely, when the surrounding AMeDAS stations are apt to fall into a basin, the \( T_{obs} \) values at the site tend to be lower than \( T_{int} \) (i.e., positive \( dT \)) because negative deviations from the climatological normal values at surrounding AMeDAS stations were interpolated to these sites. Note also that the earlier validations were conducted using AMeDAS data, which were usually of urban area, whereas most of the present sites were located near farmland. Sameshima et al. (2007) compared 40 year temperature records obtained at several AMeDAS stations and at a meteorological station surrounded by farmland and reported a greater urbanization effect at AMeDAS

---

**Fig. 5.** Deviations of the interpolated daily mean air temperatures plotted against distances between observation sites and the closest AMeDAS stations. The regression line and \( P \) value of the slope are calculated from the median values (open circles; \( N = 88 \)). Different stations are plotted in different colors. [A color version of the figure is available online]

**Fig. 6.** Deviations of the interpolated daily mean air temperatures plotted against the elevations of the observation sites. The regression line and \( P \) value of the slope are calculated from the median values (open circles; \( N = 88 \)). Different stations are plotted in different colors. [A color version of the figure is available online]

**Fig. 7.** Two-dimensional histograms of the deviations of the interpolated daily mean air temperatures plotted against (a) daily mean air temperatures and (b) daily mean wind speeds estimated by AMGSD. [A color version of the figure is available online]
stations. This could lead to greater positive deviations at AMeDAS stations and hence cause overestimations of $T_{aw}$ at interpolated sites near farmland. The climatological normal temperatures at each site might inherently contain a bias. As the normal temperatures were estimated by the multiple regression method (JMA, 2011), the possibility of the appearance of extreme estimates was lowered, as mentioned in the Introduction. In any case, the RMSE values calculated by using the AMeDAS dataset (Seino, 1993; Ohno et al., 2016) may not always reflect the value in farmlands, and the considerably greater RMSEs in the present study may reflect this.

### 3.3 Spatial extents of influential topography and geography around each site

Although the surface meteorology (i.e., daily mean air temperature and wind speed) determined the range of deviations, the $dT$ values were either negative or positive (Fig. 7). As expected from this result, a simple neural network model having the surface meteorological dataset as a predictor failed to estimate $dT$ values (data not shown). These results indicate that biases in $T_{aw}$ may not be explained by the surface meteorology in the present domain and suggest that the interactions between the boundary layer wind and topography are important to estimate $dT$. By adding multimodal datasets as predictors and selecting their ROIs appropriately, the overall RMSE could be reduced by correction with machine learning (Figs. 8a and S1). When the topographical ROI was fixed at $30 \times 30$ km, the RMSE was reduced in several combinations of ROIs for the land use predictors. Decreasing the topographical ROI to $30 \times 30$ km or smaller reduced the RMSE in some combinations irrespective of the fixed ROIs for the other predictors. Conversely, when the topographical ROI was fixed at $120 \times 120$ km, changes in the ROIs for the land uses did not result in a reduction in the RMSE (Fig. 8a). These results suggested a greater influence of the topography compared with the land use, and the influential ROI for the topography was smaller than $30 \times 30$ km. The local topography around a site might be a major factor that determines air parcel movement and the air temperature particularly on calm and cold days with large $dT$, and thereby, incorporating the small-scale topography improved the performance of the model. At most sites located in the Tokachi Plain, northern and western mountain regions were not contained in the $30 \times 30$ km ROI (Fig. 1), implying that the large-scale topography might have little influence on the air temperature at a given site in the present study domain. Fukushima et al. (2019) simulated the surface air temperature distribution in the plain by raising the elevation of the dominant wind pass, Karikachi Pass, and found that the effects of the topography were confined to a small area close to the pass (approximately $20 \times 20$ km). Furthermore, the selected ROIs were consistent with those adopted in the multiple stepwise regression method in earlier studies applied to northern Japan (Kanno, 1997; JMA, 2011). The present results and the earlier studies again suggest that the formation of the air temperature distribution in cold regions is a locally confined event.

There were several sites where the relationship between the observed and modeled $dT$ values was expressed by a linear relationship with a slope of 1 (e.g., stations 3 and 4) and where the correction could mitigate the biases although the slope did not equal 1 (e.g., stations 15, 17, and 19) (Fig. 8b). Particularly, the biases were greatly reduced at several sites where the bias was large before the correction (stations 17 and 19). Conversely, at several sites, the modeled $dT$ values were negatively correlated with the observed values (stations 1, 2, and 20) or the model responded substantially despite the observed values being $\sim 0$ (station 18), leading to increases in RMSE values by the correction. These results indicate that the model might be effective in reducing the overall RMSE, and those at certain sites, although it could increase the systematic biases in $T_{aw}$ by unnecessary corrections.
3.4 Improvement of air temperature correction by modifying machine learning models

We modified the model architecture to reduce the overall RMSE and to resolve the problematic corrections at several sites. Based on the hypothesis that the local topographical and geographical features were involved in the formation of the heterogeneous spatial distribution in $T_{\text{obs}}$, the two-dimensional datasets at smaller ROIs as well as the effective ones were simultaneously fed to the model. Furthermore, we randomly nested the network so that the model could learn the features intensively. One complicated model represented a smaller overall RMSE value ($1.08^\circ\text{C}$; Figs. 9 and S2). Since the sites where the correction worked were the same for two models (e.g., stations 8, 11, 17, and 19 in Figs. 8b and 9) and other models with different architectures and hyperparameters (data not shown), it was unlikely that the correction originated from false positives. This result may indicate that the models captured certain features related to the formation of the spatial distribution of the air temperature. However, the problematic corrections remained to be found at the same sites in the modified model and other tested models, suggesting a limited generalization ability of the machine learning techniques for the air temperature corrections. In the present study, therefore, we could not develop a model that corrected $T_{\text{in}}$ at a given site and improve the spatial distribution of the daily mean air temperatures in Hokkaido during winter. Although the reason for this cannot be presently elucidated, the corrections failed on days with a specific wind direction (Figs. 8b and 9). Since the wind direction and speed exhibit diurnal variations, adopting predictors with finer temporary resolutions may improve the model.

Although the relief in the Okhotsk region was remarkable compared with the Tokachi region (Fig. 1), the correction did not work at several sites in the Okhotsk region. Although we trained several models using only Okhotsk data, the models did not exhibit better test performance at these sites (data not shown), suggesting that the distinctive predictors did not always bring about informative features. Furthermore, this result might suggest that we should further subdivide the region to develop a custom model specialized for making air temperature corrections in a given region.

3.5 Use of machine learning techniques in meteorology: Implications from the case study

In the present study, focusing on the local heterogeneous air temperature distribution in cold and snow-covered regions, we attempted to develop a multimodal machine learning model that corrected biases in spatially interpolated air temperatures. While periodicity and autocorrelation in time series data may serve as necessary features in machine learning for prediction, there was little spatial relationship among the air temperatures recorded at the observation sites in the present study (Fig. 5). This weak spatial relationship might be an obstacle to feature extraction in machine learning and effective corrections. It may be effective to use two-dimensional time series predictors such as satellite remote sensing indices (e.g., land surface temperature). Several researchers have attempted to use satellite data to improve accuracy and/or spatial resolution of surface meteorological data (e.g., Li et al., 2018; Ruiz-Álvarez et al., 2019). Although we did not adopt meteorological satellite data in the present model because there was no compatible daily dataset that covered the observation period (2010–2019), the use of satellite datasets might be a promising solution to provide accurate and finely resolved maps of meteorological elements.

A limited amount of machine learning studies has been reported in terms of downscaling of air temperatures (Baño-Medina et al., 2019; Sachindra and Kanae, 2019), whereas considerable success has been achieved in downscaling of precipitation (e.g., Wilby et al., 1998; Vandal et al., 2017; Misra et al., 2018; Baño-Medina et al., 2019; Pan et al., 2019). Gaps between model formulas and actual events of rainfall processes may be greater than those in temperature, and hence, there may be still room for the contribution of machine learning. The impossibility of answering a task with a given dataset in machine learning cannot be proven.
This fact can be an obstacle for reporting “negative results” (Schooler, 2011) in downscaling studies of air temperature. Furthermore, urgent demands for prediction and downscaling of precipitation might encourage such attempts, thereby leading to an increase in the number of reports that exploit machine learning techniques to analyze rainfall events. Reporting negative results along with the details of the data, methods, and approaches should be promoted in machine learning studies in meteorology to avoid publication bias (Sterling, 1959).

4. Concluding remarks

In the present study, we revisited the accuracy of spatially interpolated estimates of daily mean air temperatures in the Tokachi and Okhotsk regions in eastern Hokkaido during winter, where boundary layer wind and topography generate a non-uniform air temperature distribution. The median deviations were greater than +1°C or smaller than −1°C in 14 out of 88 sites. The variance of $dT$ was greater on cold and calm days, suggesting the involvement of radiative cooling and the accumulation of cold air parcels on the $dT$ values. To estimate $dT$ by considering the formation mechanisms of the air temperature distribution, we attempted to develop a machine learning model with surface and boundary layer meteorological data and the surrounding topographical and geographical features as predictors. Multimodal machine learning by changing the ROIs for the predictors suggested that the effect of the topography around a site appeared to be greater than that of land use and revealed that an influential ROI for topography was approximately 30 × 30 km or smaller, suggesting the involvement of locally confined meteorological events. While applying a model that took only surface meteorological data failed to correct the bias, applying the multimodal models reduced the overall biases by ~0.1°C in the spatially interpolated daily mean air temperature. Therefore, our results suggest that the multimodal machine learning can be effectively used in combination with other downscaling approaches. However, the correction sometimes increased the biases of $T_{\text{ref}}$ at several sites. Further improvements, such as the use of satellite remote sensing data, are needed for the practical application of machine learning methods in local meteorology.

Author contributions

K.M. designed the study, analyzed, collected, and handled data, and wrote the manuscript. T.H. designed the study and made critical revision of the manuscript. S.S. and T.Y. conducted meteorological field observation and made critical revision of the manuscript.

Acknowledgments

Meteorological data collected in the present study were kindly provided by Tokachi Federation of Agricultural Cooperatives, JA Okhotsk Abashiri, Shari town, Town of Ozora, Bihoro Town, JA Kitamirai and Obihiro City. We thank Shinni Yokoyama (AGW), Mutsuko Suzuki (AGW), and Iwao Oshima (JWA) for supporting data migration; Hirokazu Fukushima (JMA) for helpful discussion; Shigefumi Hatakeyama (JA Kitamirai) and Yuuya Ito (JA Kitamirai) for validating meteorological observation data; Dr. Manabu Nemoto (HARC/NARO) and Dr. Satoshi Inoue (HARC/NARO) for comments on the manuscript. This work was financially supported by JSPS KAKENHI Grant Number JP19H00963.

References

Allaire J, Chollet F, 2018: Keras: R interface to ‘keras’.
Allaire J, Tang Y, 2018: Tensorflow: R interface to ‘tensorflow’.
Ashworth E, Wisniewski M, 1991: Response of fruit tree tissues to freezing temperatures, HortScience 26, 501–504.
Baño-Medina J, Manzanas R, Gutiérrez JM, 2019: Configuration and intercomparison of deep learning neural models for statistical downscaling. Geoscientific Model Development Discussions https://doi.org/10.5194/gmd-2019-278, in review.
Dong C, Loy CC, He K, Tang X, 2014: Image super-resolution using deep convolutional networks. arXiv abs/ 1501.00092.
Feidas H, Karagiannidis A, Keppas S, Vaitis M, Kontos T, Zanis P, Melas D, Anadranistakis E, 2014: Modeling and mapping temperature and precipitation climate data in Greece using topographical and geographical parameters. Theoretical and Applied Climatology 118, 133–146.
Fowler D, Limin A, Ritchie J, 1999: Low-temperature tolerance in cereals: Model and genetic interpretation. Crop Science 39, 626–633.
Fukushima H, Yazaki T, Hirota T, Iwata Y, Wajima A, Yokota A, 2019: Factors and mechanisms affecting the air temperature distribution on a clear winter night in a snow-covered mesoscale plain. Journal of the Meteorological Society of Japan 97, 105–121.
Geospatial Information Authority of Japan, 2013: A server-side program to calculate elevation from longitude and latitude. Retrieved http://maps.gsi.go.jp/development/elevation_s.html (visited on 2020/02/28).
GFD Dennou Staff, 2008: GFD-DENNOU Club, 2008: Retrieved https://www.gfd-dennou.org/index.html.en (visited on 2020/02/28).
Hirota T, Usuki K, Hayashi M, Nemoto M, Iwata Y, Yanai Y, Yazaki T, Inoue S, 2011: Soil frost control: agricultural adaptation to climate variability in a cold region of Japan. Mitigation and Adaptation Strategies for Global Change 16, 791–802.
Hirota T, Yamazaki T, Yasui Y, Furukawa J, Niwa K, Nemoto M, Hamasaki T, Shimoda S, Kanno H, Nishio Z, 2017: Climate change favors viticulture and wine production in Hokkaido, Japan.—An example of Pinot Noir cultivation since 1998—(in Japanese). Climate in Biosphere 17, 34–45.
Hirota T, Kobayashi K, 2019: The roles of farmers, scientists, and extension staff in technology development for soil frost control as an adaptation to climate change in Tokachi, Hokkaido, Japan. In Adaptation to Climate Change in Agriculture (ed. by Iizumi T, Hirata R, Matsuda R). Springer, Singapore, pp. 211–228.
Japan Aerospace Exploration Agency, 2019: ALOS world 3D–30m (ver. 2.2).
Japan Meteorological Agency, 2011: JMA’s new climatological normals for 1981–2010.
Kanne H, 1997: Estimation of daily-mean air temperatures on a 1 km mesh during the occurrence of a Yamase wind (in Japanese). Journal of Agricultural Meteorology 53, 11–19.
Kimura F, 1986: Formation mechanism of the nocturnal mesoscale vortex in Kanto Plain. Journal of the Meteorological Society of Japan. Ser. II 64, 857–870.
Kondo J, 2000: Atmospheric Science Near the Ground Surface—Understanding and Application—(in Japanese). University of...
Tokyo Press, Tokyo, pp. 109–136.
Li L, 2019: Geographically weighted machine learning and downscaling for high-resolution spatiotemporal estimations of wind speed. Remote Sensing 11, 1378; https://doi.org/10.3390/rs11111378.
Li X, Zhou Y, Asrar GR, Zhu Z, 2018: Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous united states. Remote Sensing of Environment 215, 74–84.
Miao Q, Pan B, Wang H, Hsu K, Sorooshian S, 2019: Improving monsoon precipitation prediction using combined convolutional and long short term memory neural network. Water 11, 977; https://doi.org/10.3390/w11050977.
Misra S, Sarkar S, Mitra P, 2018: Statistical downscaling of precipitation using long short-term memory recurrent neural networks. Theoretical and Applied Climatology 134, 1179–1196.
National Land Numerical Information, 2016; Land use fragmented mesh (ver. 2.6). Ministry of Land, Infrastructure, Transport and Tourism.
Nemani R, Hirota T, Sato T, 2016: Prediction of climatic suitability for wine grape production under the climatic change in Hokkaido. Journal of Agricultural Meteorology 72, 167–172.
Ninyerola M, Pons X, Roure JM, 2000: A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. International Journal of Climatology 20, 1823–1841.
Ohno H, Sasaki K, Ohara G, Nakazono K, 2016: Development of grid square air temperature and precipitation data compiled from observed, forecasted, and climatic normal data (in Japanese). Climate in Biosphere 16, 71–79.
Ohno H, Sasaki K, 2019: 1-km grid meteorological data service and its use to reduce weather and climate risks in field crop production. In Climate smart agriculture for the small-scale farmers in the Asian and Pacific region (ed. by Shirato Y., Hasebe A.). Institute for Agro-Environmental Sciences, NARO, Japan, pp. 313–328.
Pan B, Hsu K, AghaKouchak A, Sorooshian S, 2019: Improving precipitation estimation using convolutional neural network. Water Resource Research 55, 2301–2321.
Pardo-Iguzquiza E, 1998: Comparison of geostatistical methods for estimating the areal average climatological rainfall mean using data on precipitation and topography. International Journal of Climatology 18, 1031–1047.
Pielke SRA, Wilby RL, 2012: Regional climate downscaling: What’s the point? Eos, Transactions American Geophysical Union 93, 52–53.
R Core Team, 2019: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Revelle W, 2018: Psych: Procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois.
Ruiz-Álvarez M, Alonso-Sarria F, Gomariz-Castillo F, 2019: Interpolation of instantaneous air temperature using geographical and MODIS derived variables with machine learning techniques. ISPRS International Journal of Geo-Information 8, 382; https://doi.org/10.3390/ijgi8090382.
Sachindra D, Kanae S, 2019: Machine learning for downscaling: The use of parallel multiple populations in genetic programming. Stochastic Environmental Research and Risk Assessment 33, 1497–1533.
Sameshima R, Hirota T, Hamasaki T, Suzuki S, 2007: Temperature trends at the National Agricultural Research Center for Hokkaido Region in the 40 years from 1966 to 2005. Journal of Agricultural Meteorology 63, 95–102.
Sameshima R, Yokoyama S, Hirota T, Hamasaki T, 2008: Creation and application of 250 m square grid meteorological information for crop management using a local weather station network. Japan Agricultural Research Quarterly 42, 41–48.
Schooler J, 2011: Unpublished results hide the decline effect. Nature 470, 437.
Seino H, 1993: An estimation of distribution of meteorological elements using GIS and AMeDAS data (in Japanese). Journal of Agricultural Meteorology 48, 379–383.
Shi X, Chen Z, Wang H, Yeung D, Wong W, Woo W, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. arXiv abs/1506.04214.
Shimoda S, Yazaki T, Nishio Z, Hamasaki T, Hirota T, 2015: Possible soil frost control by snow compaction on winter wheat fields. Journal of Agricultural Meteorology 71, 276–281.
Sterling TD, 1959: Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. Journal of the American Statistical Association 54, 30–34.
Ueyama H, 2008: Cartography of 50m grid climate map of monthly mean air temperature and its utilization (in Japanese). The Journal of the Japanese Agricultural Systems Society 24, 191–198.
Ueyama H, 2013: Classification of recent studies by method type for surface air temperature map development and estimation of daily temperature using a radiative cooling scale. Journal of Agricultural Meteorology 69, 215–227.
Vandal T, Kodra E, Ganguly S, Michaelis AR, Nemani RR, Ganguly AR, 2017: DeepSD: Generating high resolution climate change projections through single image super-resolution. arXiv abs/1703.03126.
Wilby RL, Wigley T, Conway D, Jones P, Hewitson B, Main J, Wilks DS, 1998: Statistical downscaling of general circulation model output: A comparison of methods. Water Resource Research 34, 2995–3008.
Yazaki T, Hirota T, Iwata Y, Inoue S, Usuki K, Suzuki T, Shirahata M, Iwasaki A, Kajiyama T, Araki K, Takemiy Y, Maezuka K, 2013: Effective killing of volunteer potato (Solanum tuberosum L.) tubers by soil frost control using agrometeorological information—An adaptive countermeasure to climate change in a cold region. Agricultural and Forest Meteorology 182, 91–100.
Yazaki T, Fukushima H, Hirota T, Iwata Y, Wajima A, Yokota A, 2017: Winter nocturnal air temperature distribution for a mesoscale plain of a snow-covered region: Field meteorological observations and numerical simulations. Journal of Applied Meteorology and Climatology 56, 519–533.
Yazaki T, Hirota T, 2019: Controlling the depth of soil frost in farm fields in Japan. In Adaptation to Climate Change in Agriculture (ed. by Izumi T, Hirota R, Matsuda R). Springer, Singapore, pp. 53–65.
Yonekura K, Hattori H, Suzuki T, 2018: Short-term local weather forecast using dense weather station by deep neural network. 2018 IEEE International Conference on Big Data (BigData 2018), December 10–13, Seattle, pp. 1683–1690.