Sensors and Materials, Vol. 38, No. 9 (2026) 5153–5174
MYU Tokyo

S & M 4615

Microclimate Air Temperature Prediction Method Based on Difference Prediction and Transfer Learning

Kiyoto Kimura[1] and Takuya Yoshihiro[2*]

[1]Graduate School of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama City, Wakayama 640-8510, Japan
[2]Faculty of Systems Engineering, Wakayama University,
930 Sakaedani, Wakayama City, Wakayama 640-8510, Japan

(Received May 14, 2026; accepted August 10, 2026)

Keywords: microclimate prediction, deep learning, transfer learning

Microclimate air temperature varies substantially over short distances because of local factors such as buildings, vegetation, solar radiation, and shading. Dense observations can capture this variability, but long-term deployment at many target sites is costly. This study evaluates an long short-term memory (LSTM)-based framework for air-temperature prediction at data-scarce microclimate sites using a nearby public meteorological station as a reference. The model predicts the temperature difference between the target site and the reference station instead of directly predicting the target-site temperature. Transfer learning is also introduced by pretraining the model on other nearby microclimate sites and fine-tuning it using limited target-site data. The framework was evaluated using observations from Wakayama University and the automated meteorological data acquisition system (AMeDAS) Wakayama station. Difference prediction reduced root mean square error (*RMSE*) compared with direct prediction at all evaluated sites. Transfer learning did not always improve the best accuracy when sufficient target-site data were available, but it improved robustness when fine-tuning data were limited or seasonally different from the evaluation period. These results suggest that difference prediction and transfer learning can support practical air-temperature prediction at data-scarce microclimate sites under conditions similar to those examined in this study.

1. Introduction

Microclimate refers to localized weather conditions near the ground surface that are strongly affected by surrounding environments such as buildings, vegetation, solar radiation, shading, and human activities. Because these local factors vary over short distances, microclimate can differ substantially even between nearby locations.[(1)] An accurate understanding of microclimate is important for urban environmental management, heat-risk assessment, and local-scale decision making.[(2,3)]

The most reliable way to understand microclimate is dense *in situ* observation. However, high-density sensor deployment over large area and over long periods results in high cost in

*Corresponding author: e-mail: tac@wakayama-u.ac.jp
https://doi.org/10.18494/SAM6437

ISSN 0914-4935
https://myukk.org/

practice. As a result, methods are needed to predict microclimate even when the volume of observations at a target site is limited.

Recent studies on microclimate prediction have explored several directions. One major approach is downscaling from coarse regional data such as numerical weather prediction (NWP) data by incorporating urban morphology, land-cover information, or three-dimensional city models.(4–6) Another approach uses data collected from dense local sensor networks and applies deep learning to learn spatiotemporal relationships for future microclimate prediction.(7,8) These methods have demonstrated high prediction performance. However, they often require detailed auxiliary data from multiple sources, dense sensor networks, or long observation periods. These characteristics raise two practical issues for microclimate prediction. First, the amount of data available at a target site is often insufficient for training deep learning models. Second, many existing methods depend on complex input data that may not be readily available in practical applications.

Several studies have also investigated weather-station-based or location-transfer approaches for microclimate and urban air-temperature prediction. DeepMC predicts the error between weather-station forecasts and local microclimate observations using a multi-scale encoder-decoder deep learning framework.(7) Yang *et al.* predicted urban microclimate conditions from local weather-station data using an artificial neural network.(9) Shen proposed a reference-station-based machine learning framework for the spatiotemporal mapping of urban air temperature and urban heat island patterns.(10) In addition, Li *et al.* applied deep transfer learning to microclimate estimation using sensor readings and building morphological features,(11) and Deznabi *et al.* investigated zero-shot microclimate prediction for new or unmonitored locations.(12)

These studies show the usefulness of weather-station-based prediction, reference-station-based air-temperature estimation, and knowledge transfer for microclimate prediction. However, it remains insufficiently examined whether predicting the temperature difference from a reliable nearby public meteorological station improves microclimate air-temperature prediction compared with directly predicting the target-site temperature. It also remains unclear whether pretraining on neighboring microclimate sites and fine-tuning with limited target-site data improves robustness when the amount or season of the fine-tuning data is constrained.

In this study, we predict air temperature at a target microclimate site from observations at a nearby public reference station such as Automated Meteorological Data Acquisition System (AMeDAS).(13) Instead of directly predicting the target-site temperature, we train a deep learning model to estimate the temperature difference between the target site and the reference station. The final target-site temperature is reconstructed by adding the predicted difference value to the observed temperature at the reference station. This design is motivated by the idea that large-scale temporal variation is already captured by the reference station, while the remaining difference mainly reflects local deviations caused by the surrounding environment. Therefore, predicting the difference can simplify the learning target and stabilize training.

We further introduce transfer learning to improve robustness when the amount of target-site data is limited. The model is first pretrained using data from other nearby microclimate sites and is then fine-tuned for the target site. This framework is intended to support practical deployment at locations where only a short observation period is available.

This study does not aim to propose a new forecasting architecture. Instead, it empirically evaluates a controlled long short-term memory (LSTM)-based framework for data-limited microclimate air-temperature prediction. Specifically, this study has three objectives. First, we evaluate whether difference prediction improves prediction accuracy compared with direct prediction under the same LSTM architecture. Second, we examine whether transfer learning improves robustness when the amount of target-site data is limited or when the fine-tuning data are collected in a season different from the evaluation period. Third, we analyze the effects of site-specific auxiliary information and source-site selection on transfer performance. Through these evaluations, this study clarifies the practical effectiveness and limitations of combining difference prediction and transfer learning for microclimate air-temperature prediction under realistic observational constraints.

This paper is organized as follows. In Sect. 2, we describe the proposed method. In Sect. 3, we provide evaluation methods and results to show the performance of the proposed method. After giving a discussion in Sect. 4, we finally conclude the work in Sect. 5.

2. Proposed Method

2.1 Overview of the prediction method

The objective of this study is to predict microclimate from the observations at the nearest public station by using deep learning. Also, we assume a practical setting in which newly installed microclimate observation sites have only a limited observation period. Under this setting, we examine whether a transfer learning-based method can help maintain prediction performance when only a small amount of data is available at target sites.

In this study, we focus on temperature prediction. Figure 1 shows the overall framework of the proposed method. First, we install a sensor at the target microclimate observation point and measure temperature for a certain period. And, we identify the nearest public meteorological station to the target site as the reference station. We then use deep learning to learn the relationship between the temperature observed at the reference station and that observed at the target site.

A key feature of the proposed method is that the model does not directly predict the temperature at the microclimate site, y_{micro}. Instead, the model predicts the difference value y_{diff}, which is defined as the residual from the reference station temperature y_{ref}, i.e., $y_{diff} = y_{micro} - y_{ref}$.

During inference, we assume that only the observation at the reference station, y_{ref}, is available. The trained deep learning model estimates the difference value $\hat{y}_{diff}$. We then compute the temperature at the microclimate site by using Eq. (1), as shown in the lower part of Fig. 1.

$$\hat{y}_{micro} = y_{diff} + y_{ref} \tag{1}$$

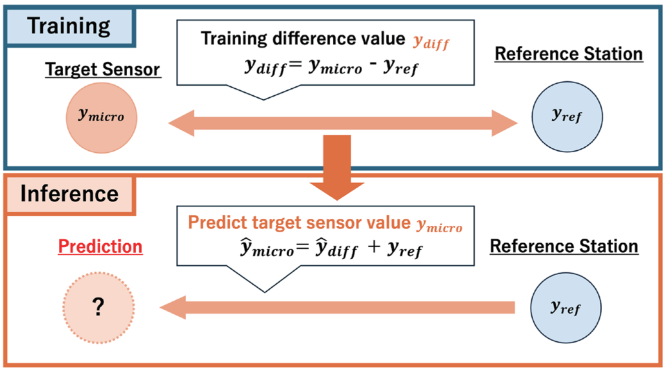


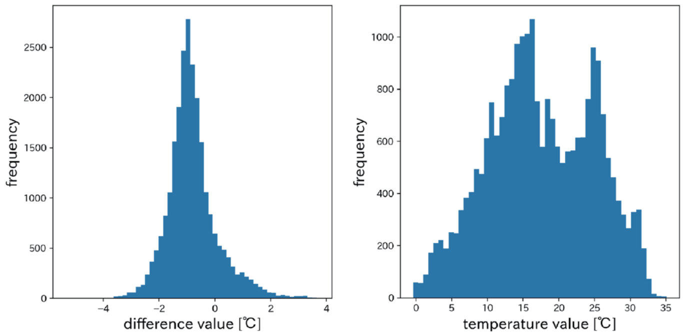
Fig. 1. (Color online) Overview of the proposed prediction method.

Learning the difference values has an important advantage. By subtracting the large-scale temperature variation represented by the reference station in advance, the deep learning model can focus on estimating the deviation caused by the local environment. As a result, the range and distribution of the target variable become relatively simple. This property is expected to stabilize model training.

Figure 2 shows histograms of the temperature values and the corresponding difference values at one microclimate site over approximately one year. The left figure shows the histogram of the difference values and the right figure shows the histogram of the temperature values. In both figures, the horizontal axis represents the variable value and the vertical axis represents the frequency. As shown in Fig. 2, the temperature at the microclimate site is distributed over a wide range, whereas the difference values have a narrower distribution and are more concentrated around the center.

2.2 Prediction model

Here we describe the prediction model used to estimate the temperature difference value y_{diff} at the target observation site. In this study, we use a prediction model based on LSTM,[14] which is well suited to time-series forecasting. Figure 3 shows the architecture of the prediction model. The input variables are meteorological observations obtained from the reference station. The output variable is the temperature difference value y_{diff} between the target site and the reference station. Each input sequence consists of data from the prediction time back to 24 h before that time. The LSTM layer receives these sequential inputs and processes 24 h of data with 144 time steps at a 10 min resolution. The model then converts the extracted temporal features into the predicted temperature difference value. We use a single-layer LSTM with 100 hidden units. We then place two fully connected layers after the LSTM layer. The first fully connected layer has 50 units and the second fully connected layer has 10 units. Finally, the output layer produces a single predicted value of y_{diff}.



(a) (b)

Fig. 2. (Color online) Histograms of difference values (a) and temperature values (b) at a microclimate site (Site 1).

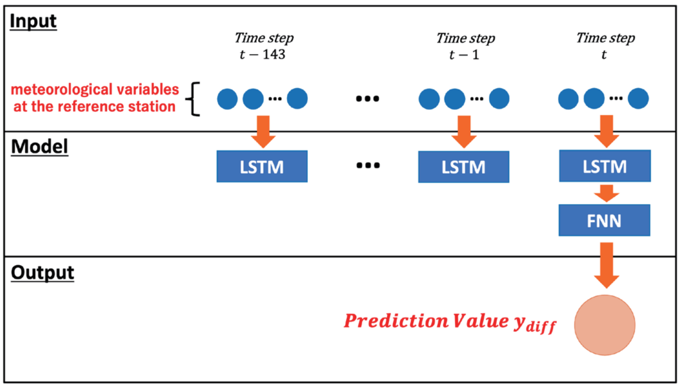


Fig. 3. (Color online) Architecture of the prediction model based on LSTM.

2.3 Prediction with transfer learning

2.3.1 Introduction of site-specific feature and generation of its estimated value

In this subsection, we introduce illuminance as a site-specific auxiliary feature for the difference value prediction model described in Sect. 2.1. Microscale temperature variations are affected by local environmental factors such as building layout, solar radiation, shadows, and surface conditions. Therefore, when only meteorological variables observed at the reference station are used as inputs, the model may not sufficiently represent site-specific variations. In this study, we use illuminance as an auxiliary feature because it reflects the effects of solar radiation and shading, which are important factors causing inter-site differences in microclimate temperature.

In practical operation, however, we cannot always observe the site-specific feature continuously at the target site. If a continuous observation of the site-specific feature is required, we must permanently install an additional sensor at the target site. This requirement is inconsistent with the intended operational scenario, in which we estimate microscale meteorological conditions from the observations at the reference station. Therefore, in this study, we observe the site-specific feature only for a limited period at the target site. We then generate an estimated site-specific feature from that limited observation and use it as an input during inference. Below, we describe the estimation procedure when illuminance is adopted as the site-specific feature.

In the following formulation, $L_{ref}(d, t)$ denotes the illuminance corresponding to the reference area. Ideally, this value is observed at the public reference station itself. However, some public meteorological stations, including the AMeDAS Wakayama station used in this study, do not provide illuminance. In such cases, illuminance observed at a nearby open location can be used as a proxy for the reference-area illuminance, provided that the location is not strongly affected by local shadows and represents the large-scale solar condition around the reference station. Therefore, $L_{ref}(d, t)$ in Eqs. (2) and (3) can be interpreted either as illuminance directly observed at the reference station or as proxy illuminance observed at a nearby representative location.

Let $L_{ref}(d, t)$ denote the illuminance observed at the reference station and let $L_{tar}(d, t)$ denote the illuminance observed at the target site. Here, d denotes the day and t denotes the time. To generate the estimated illuminance, we use the 15 days immediately before the prediction period as a calibration period. During the calibration period, we judge a time slot (d, t) as sunny when the sunshine duration observed at the reference station (AMeDAS) is 600 s at each 10 min interval (d, t). We then define $\mathcal{D}_{sun}(t)$ as the set of days for which time slot t is judged as sunny. For example, if time slot t is judged as sunny on 10 out of the 15 calibration days, then $|\mathcal{D}_{sun}(t)| = 10$.

Next, we introduce a time-dependent illuminance ratio $p(t)$ for each time slot t. When $|\mathcal{D}_{sun}(t)| > 0$, we define $p(t)$ as the average illuminance ratio over sunny days, as follows.

$$p(t) = \frac{1}{|\mathcal{D}_{sun}(t)|} \sum_{d \in \mathcal{D}_{sun}(t)} \frac{L_{tar}(d,t)}{L_{ref}(d,t)} \tag{2}$$

In contrast, when no sunny sample exists for time slot t, for example during night time, we set $p(t) = 1$. Using this ratio, we estimate the illuminance at the target site during the prediction period by

$$\hat{L}_{tar}(d,t) = p(t) L_{ref}(d,t). \tag{3}$$

This method assumes that the shading environment does not change between the calibration period and the prediction period. In other words, we assume that buildings, trees, and sensor locations remain fixed. Under this assumption, we regard the estimated illuminance ratio $p(t)$ as applicable to the prediction period. We then input the generated $\hat{L}_{tar}(d, t)$ to the model together with the meteorological variables obtained from the reference station and estimate the difference value y_{diff}.

2.3.2 Transfer learning of the prediction model

In this study, we apply transfer learning to stabilize training and maintain prediction accuracy at target sites with limited data. The purpose of transfer learning is to share knowledge obtained from multiple microclimate sites and reduce the amount of training data required for adaptation to a new site. We perform transfer learning in two stages: pretraining and fine-tuning.

In the pretraining stage, we train the difference value prediction model shown in Fig. 3 by using the temperature data from microclimate sites other than the target site. At this stage, we use both the meteorological variables at the reference station and the illuminance corresponding to each site as inputs. We use the difference value y_{diff} as the output variable and optimize the model parameters. Through pretraining, we enable the model to acquire general temporal patterns of meteorological variation and to learn how to incorporate the effects of the local environment through the use of illuminance.

Next, in the fine-tuning stage, we retrain the model by using the temperature data at the target site and adapt the model to the site-specific characteristics. In this study, we update only the parameters of the fully connected layers during fine-tuning and keep the LSTM layer fixed. We make this design choice because the LSTM layer extracts features with high commonality across sites, such as diurnal cycles and air mass changes. In contrast, the output layers map the learned representation to the difference value and therefore play a central role in site-specific prediction. Updating all layers with limited fine-tuning data can easily cause overfitting. Restricting the updated parameters to the output layers reduces the degree of freedom of the model and enables stable adaptation even under limited-data conditions.

2.4 Training settings of the prediction model

In this subsection, we describe the training settings of the deep learning model for microscale meteorological prediction. Table 1 shows the settings used in this study. We use Adam as the optimizer. We change the initial learning rate between the two stages of transfer learning, namely, pretraining and fine-tuning. Specifically, we set the initial learning rate to 0.001 for pretraining and 0.0001 for fine-tuning. Fine-tuning uses a smaller amount of training data and aims to adjust only the trainable parameters in the fully connected layers. Therefore, we use a learning rate lower than that used in pretraining to stabilize learning.

Table 1
Training settings of the prediction model

Item	Setting
Optimizer	Pretraining: Adam (initial learning rate: 0.001) Fine-tuning: Adam (initial learning rate: 0.0001)
Stopping criterion	Early stopping (patience = 10)
Loss function	*MSE* (mean squared error)
Batch size	128
Language	Python version 3.12.4
Library	PyTorch version 2.3.1

3. Results

3.1 Evaluation method

In this section, we design a deep-learning-based method for microclimate prediction and evaluate whether the method can achieve reasonable accuracy under the limited observational resources considered in this study. The prediction task is to estimate temperature at microclimate sites. We use the seven microclimate sites in Wakayama University shown in Fig. 4. We use the Wakayama station of AMeDAS as the reference station. We define the temperature difference between the reference station and each target sensor as the prediction target. We then estimate the target site temperature using the predicted difference.

We evaluate the method from two perspectives. First, we evaluate the accuracy of the difference prediction model. In this evaluation, we compare the difference prediction model without transfer learning with a model that directly predicts the target site temperature. Second, we evaluate the effect of transfer learning. In particular, we examine the prediction accuracy with transfer learning and its robustness when the amount of fine-tuning data is reduced.

In the evaluation, we use six sites as prediction targets (Sites 1–5 and Site 7). We exclude Site 6 because it plays the special role of a proxy for illuminance at the reference station. This role differs from that of the other sites.

We conduct the evaluation by cross-validation. We divide the available observation period into K = 10 temporally ordered folds. For the evaluation of the difference-prediction model without transfer learning, we use leave-one-fold-out cross-validation. We use one fold for evaluation and the remaining folds for training. This setting allows us to assess the generalization performance over the full year.

For the evaluation of transfer learning, we limit the evaluation period to two seasons, namely, winter and summer. In principle, we use data outside the evaluation period for pretraining. We vary the period used for fine-tuning in accordance with each experimental setting. This design allows us to compare performance under reduced-data conditions. We use *RMSE* as the evaluation metric.

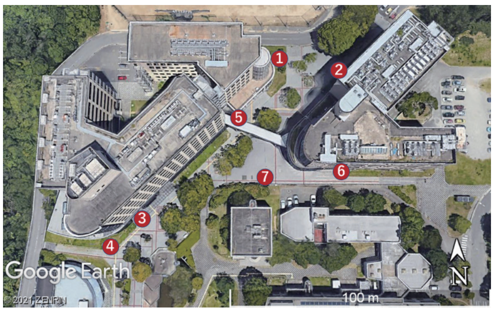


Fig. 4. (Color online) Microclimate observation sites (source: Google Earth).

3.2 Data

3.2.1 Data acquisition

Figure 4 shows the microclimate observation sites used in this study. These seven sites are located around North Building No. 1 on the Wakayama University campus. At each site, we installed an OMRON environmental sensor (2JCIE-BL01, Fig. 5) inside a ventilation duct. The sensors recorded temperature, relative humidity, illuminance, and atmospheric pressure at 1 min intervals. The observation period lasted from February 4, 2021 to December 2, 2021.

We preprocess the collected microclimate data before model training. The preprocessing includes removing time points with missing values and converting the temporal resolution. We use 10 min resolution time series data for learning because 1 min resolution data are more sensitive to missing values and short-term noise.

We use the Wakayama AMeDAS station, which is the nearest public weather station, as the reference station. From this station, we obtain temperature, precipitation, wind speed, wind direction, and humidity. We use these variables as input features in the deep learning model. We also add time-of-day and date information as input features.

The AMeDAS Wakayama station does not provide illuminance. Therefore, in this experiment, we used the illuminance observed at Site 6 in Fig. 4 as a proxy for the reference-area illuminance. Site 6 was selected because it was located in a relatively open area and was less affected by shadows from surrounding buildings than the other microclimate sites. Since all observation sites were located within the same campus-scale area, we regarded the illuminance at Site 6 as an approximate measure of the large-scale solar condition shared by the reference area and the target sites. This proxy was used only because illuminance was unavailable from the public reference station; if illuminance is observed at the reference station or at another nearby representative open location, such data can be used instead.

As part of preprocessing, we transform wind direction, time, and date into sine and cosine components because these variables have cyclic properties. We also apply min–max normalization to the input variables.

Fig. 5. (Color online) Microclimate observation device.

3.2.2 Dataset

After preprocessing the data from the observation period between February 4, 2021 and December 2, 2021, we obtained 156 days for which data were simultaneously available at all seven sites. We divide these 156 days into 10 temporally ordered subsets on a daily basis for training and evaluation.

The inputs to the difference prediction model consist of the meteorological variables observed at the reference station, namely, temperature, precipitation, wind speed, wind direction, and humidity, together with time and date information. In the transfer learning setting, we additionally use illuminance observed at each microclimate site as an input feature. However, in practical operation, illuminance at the target site may not be available. We therefore use estimated illuminance as the input at the inference time. In the evaluation presented in this section, we also use estimated illuminance when making predictions for the validation period.

3.3 Evaluation of the difference prediction model

In this section, we evaluate the effectiveness of the proposed method using difference prediction under the condition without transfer learning. The proposed method uses meteorological observations at the reference station, i.e., the Wakayama station, as inputs. It estimates the temperature difference at the target site and reconstructs the target-site temperature by adding the predicted difference to the reference-station temperature. In contrast, the comparison method directly predicts the target site temperature from the same inputs.

We divide the data into 10 subsets on a daily basis and conduct leave-one-fold-out cross validation. For each target site, we train an independent prediction model. We do not apply transfer learning in this experiment.

Figure 6 shows the *RMSE* values obtained for all six target sites. In this figure, the horizontal axis represents the target site and the vertical axis represents *RMSE* (°C). The bar chart shows the results for the proposed method (Proposed) and the comparison method (Conventional) that

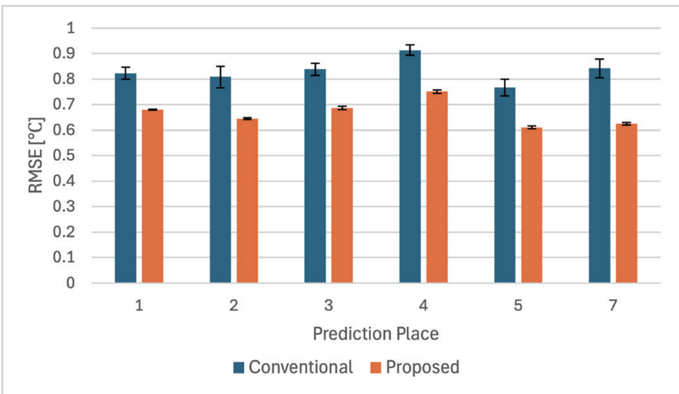


Fig. 6. (Color online) Prediction accuracy of temperature at each microclimate site.

directly predicts temperature. Figure 6 shows that the proposed method reduces *RMSE* at all sites. The *t*-test results from 10 repetitions at each of the prediction places 1–5 and 7 were 2.47 × 10^{-4}, 1.41 × 10^{-4}, 1.16 × 10^{-4}, 2.85 × 10^{-4}, 5.06 × 10^{-4}, and 2.56 × 10^{-4}, respectively, indicating sufficient significance. This result confirms that temperature estimation based on difference prediction is consistently effective. A possible reason is that the temperature difference between the reference station and the target site has a smaller variation range than the absolute temperature at the target site. This makes the prediction task by deep learning easier.

Next, Fig. 7 shows the hourly *RMSE* at each target site. This figure presents the results for Sites 1–5 and Site 7. For each site, the horizontal axis represents the time of day and the vertical axis represents *RMSE* (°C). The line plots show *RMSE* values for the proposed and comparison methods. Figure 7 shows that the proposed method outperforms the comparison method at all sites and at all times of day. This result indicates that estimation based on difference prediction

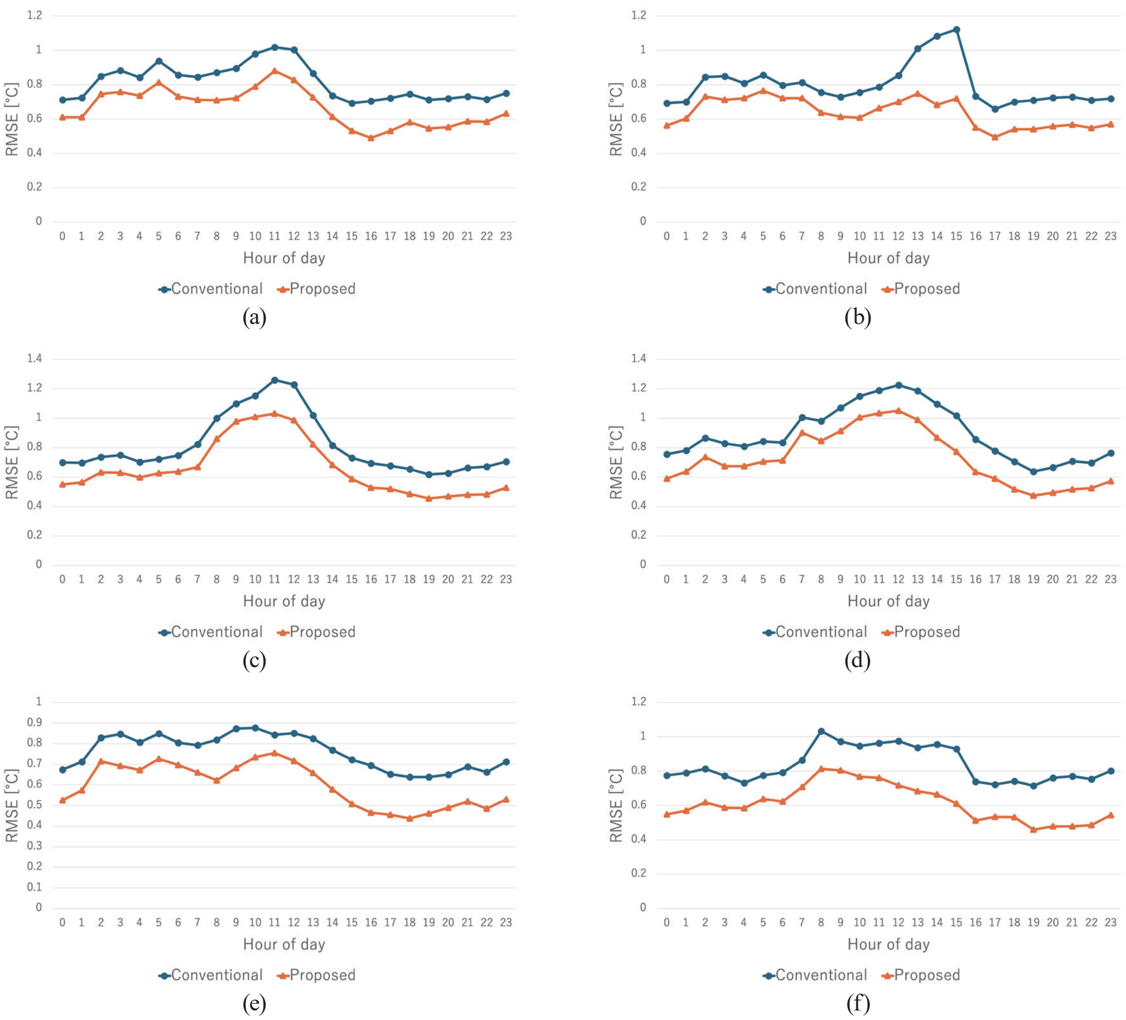


Fig. 7. (Color online) Hourly *RMSE* at each target site. (a) Site 1, (b) Site 2, (c) Site 3, (d) Site 4, (e) Site 5, and (f) Site 7.

is effective regardless of time period.

At the same time, both methods show larger errors from morning to early afternoon and smaller errors during night time. This tendency suggests that local temperature variations become stronger during the daytime because of solar radiation and surface heating. Observations from the reference station alone may not fully capture such site-specific variations.

Figure 8 shows the monthly *RMSE* at each target site. This figure shows the *RMSE* calculated for each month for the proposed and comparison methods. Figure 8 shows that the proposed method yields lower *RMSE* in most months. This result confirms that difference-prediction-based estimation is effective throughout the year.

The comparison method shows large month-to-month variation in prediction error. In contrast, the proposed method shows smaller monthly fluctuation and more stable accuracy over the year, suggesting that temperature difference prediction is more robust to seasonal changes than direct absolute temperature prediction at the target site. Moreover, the *RMSE* of the

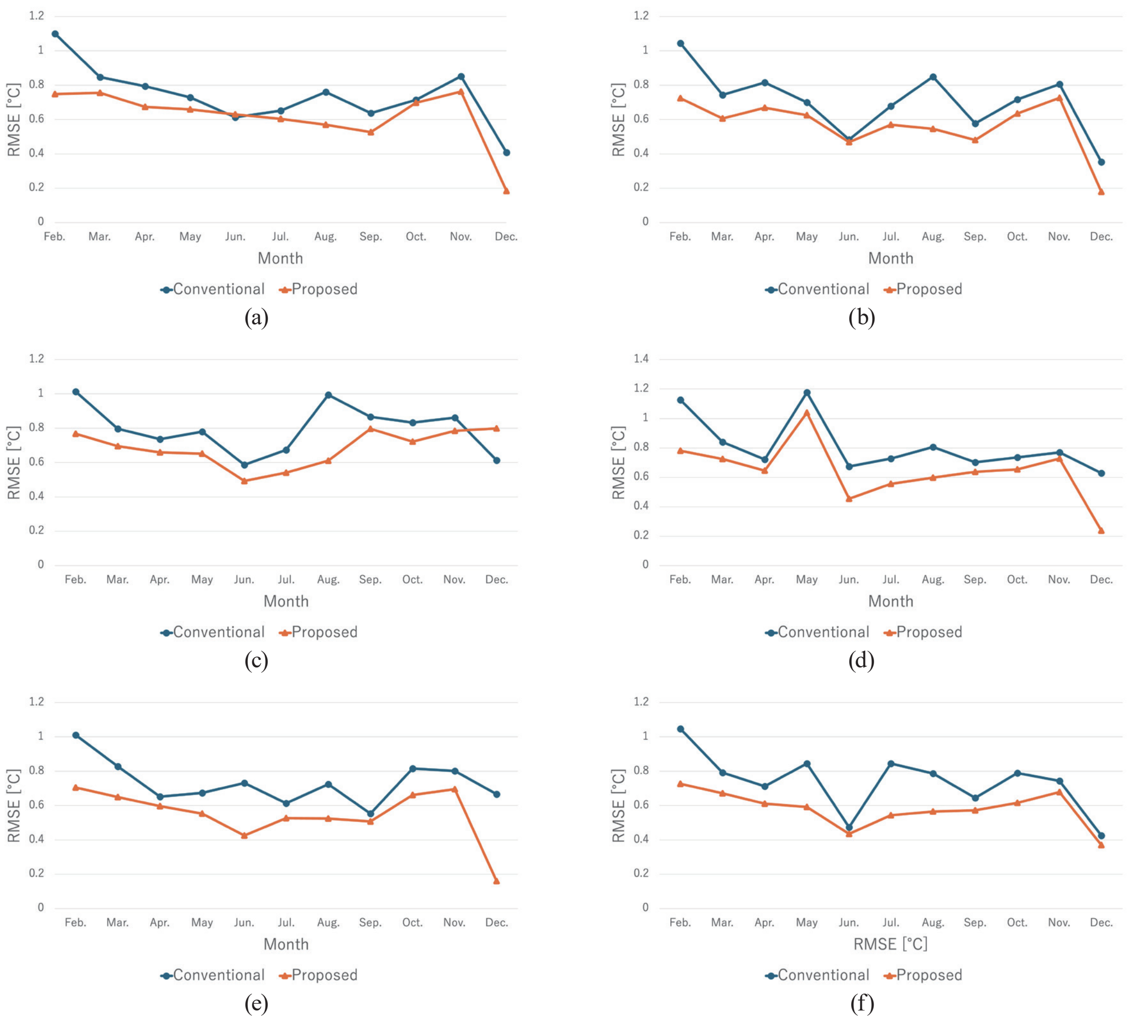


Fig. 8. (Color online) Monthly *RMSE* at each target site. (a) Site 1, (b) Site 2, (c) Site 3, (d) Site 4, (e) Site 5, and (f) Site 7.

proposed method remains approximately below 1 °C for all sites and all months. These results indicate that, in the present dataset, the method achieved *RMSE* values of approximately below 1 °C in most cases using meteorological observations from the reference station. The practical acceptability of this accuracy depends on the target application and should be examined in future deployments.

3.4 Evaluation of transfer learning

3.4.1 Prediction accuracy with transfer learning

In this section, we evaluate the prediction performance of the proposed method with transfer learning. Our goal is to develop a method that can predict microclimate with reasonable accuracy even with limited observational resources. In this study, transfer learning consists of two stages. First, we pretrain the model using data from microclimate sites other than the target site. Second, we fine-tune the model using data from the target site.

We evaluate the method in two periods of different seasons, namely, winter and summer. The winter evaluation period is 16 days from February 5, 2021, to February 20, 2021. The summer evaluation period is 16 days from July 5, 2021 to July 23, 2021.

The comparison method is the difference prediction model without transfer learning introduced in Sect. 3.3. This model estimates the temperature difference using only data from the target site and reconstructs the temperature value. The proposed method is the difference-prediction model with transfer learning. The main difference is that it incorporates data from other sites during pretraining. In this evaluation, we intentionally use the same LSTM-based architecture for both the comparison and proposed methods. The purpose of this study is not to propose a new forecasting architecture or to conduct an exhaustive benchmark among different model families. Rather, we aim to isolate the effects of the prediction formulation and the learning procedure, namely, direct training on the target site versus pretraining with non-target sites followed by fine-tuning on the target site. Therefore, changing the forecasting architecture would make it difficult to determine whether the observed performance differences are caused by transfer learning or by the model structure itself. We use LSTM as a representative and widely used deep-learning model for sequential meteorological data because the input consists of 24 h time-series sequences at 10 min resolution. Under this fixed-architecture setting, the comparison method serves as a controlled baseline for evaluating the effect of transfer learning.

We first examine the effect of transfer learning under the condition that as much training data as possible are used. For pretraining, we use data from the six non-target sites and from periods outside the evaluation period. For fine-tuning, we use data from the target site. In both stages, we exclude data belonging to the winter or summer evaluation period. In other words, we use all available data outside the evaluation period for training.

Figure 9(a) shows *RMSE* for winter prediction, and Fig. 9(b) shows *RMSE* for summer prediction. These results show that transfer learning does not produce a uniformly large improvement in prediction accuracy. It improves the accuracy at some sites, but it degrades the accuracy at others.

The difference between the proposed and comparison methods is whether data from other

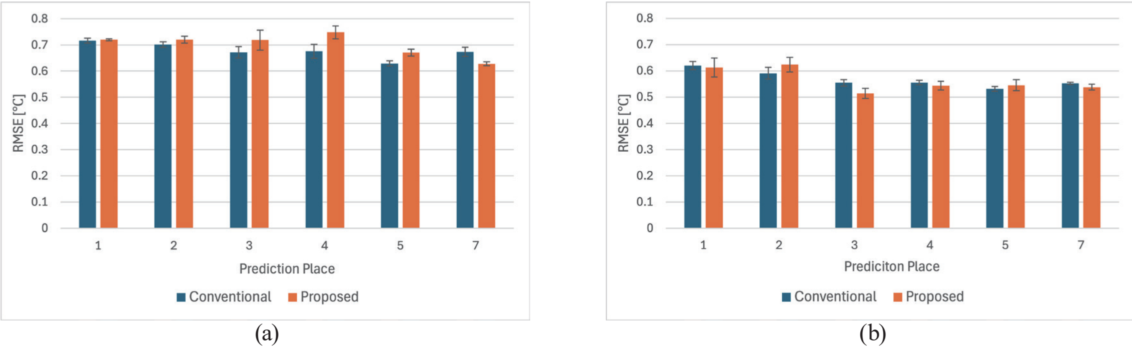


Fig. 9. (Color online) *RMSE* of temperature prediction at each site with transfer learning. (a) Winter and (b) Summer.

sites are incorporated through transfer learning. Therefore, at the sites where accuracy decreases, the pretrained representation may interfere with feature extraction that is suitable for the target site. One possible reason is that the input features do not fully capture the differences in temperature variation between the target site and the other sites.

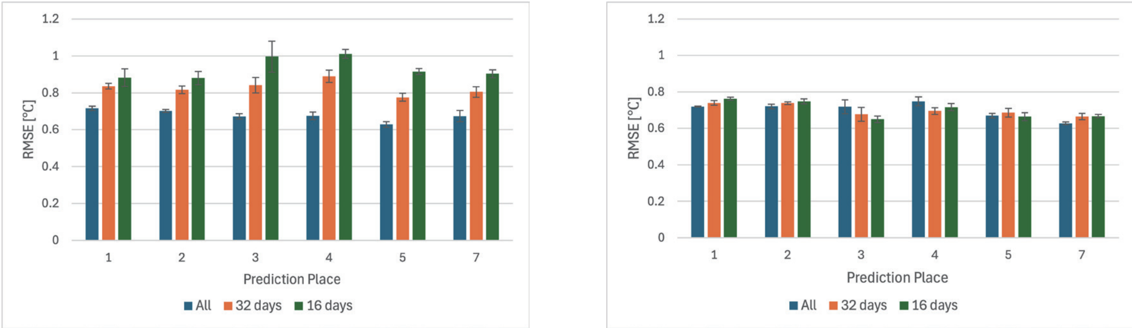
To address this issue, we include illuminance as a site-specific input feature. We expect illuminance to capture intersite differences caused by solar radiation and shading. However, factors other than solar radiation also affect microclimate. Examples include wind conditions and surface materials. If we can incorporate these site-specific factors as additional inputs, transfer from other sites may function more appropriately and further improve prediction accuracy.

3.4.2 Prediction accuracy under reduced data availability

Next, we assume a situation in which only a small amount of target-site data is available. We then evaluate whether transfer learning can maintain prediction accuracy under this condition. We use the same pretraining data as described in Sect. 3.4.1. For fine-tuning, we test three conditions for the amount of target-site data: all available periods outside the evaluation period, the most recent 32 days, and the most recent 16 days.

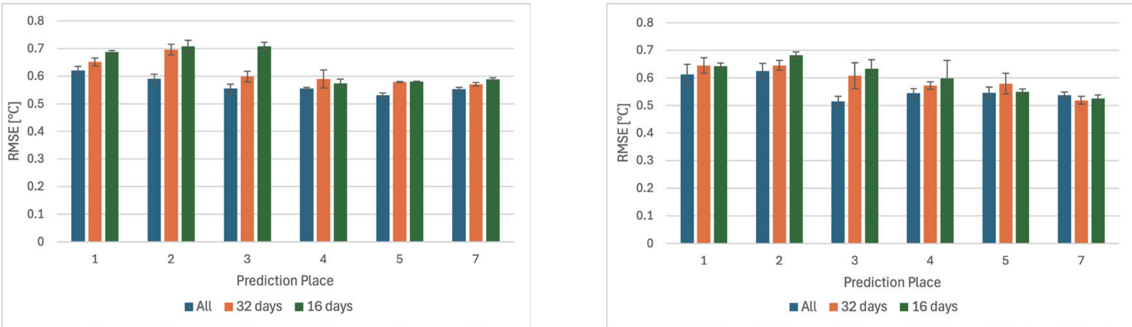
Figure 10 shows the results for winter prediction under reduced data availability. Figure 10(a) shows the *RMSE* of the comparison method and Fig. 10(b) shows the *RMSE* of the proposed method. For winter prediction, the proposed method shows only a small change in *RMSE* even when we reduce the amount of fine-tuning data from the full period to 16 days. For all places with 32- and 16-day cases, a significant difference is observed in a *t*-test at the 1% significance level, whereas no significant difference is observed for other cases. This result indicates that the proposed method can maintain prediction accuracy even with limited data. In contrast, the comparison method shows a clear increase in *RMSE* as the amount of data decreases. Its performance becomes unstable under limited-data conditions.

Figure 11 shows the results for summer prediction. For summer prediction, the proposed



(a) (b)

Fig. 10. (Color online) *RMSE* values obtained under reduced data conditions for winter prediction. (a) Comparison method (without transfer learning) and (b) proposed method (with transfer learning).



(a) (b)

Fig. 11. (Color online) *RMSE* values obtained under reduced data conditions for summer prediction. (a) Comparison method (without transfer learning) and (b) proposed method (with transfer learning).

method shows some degradation at a few sites as the amount of data decreases. However, the degree of degradation is smaller than that of the comparison method. The proposed method therefore provides relatively stable prediction even under limited-data conditions. Note that, at places 2 and 7 for the 32-day case, and places 1, 3, 5, and 7 for the 16-day case, a significant difference in a *t*-test with a significance level of 1% is observed. Overall, these results show that transfer learning enables prediction with reasonable accuracy even when the amount of target-site data is limited.

3.4.3 Prediction accuracy with fine-tuning data from different seasons

In the previous evaluations, we limited the fine-tuning data to the period immediately preceding the evaluation period. In this subsection, we assume a situation in which such recent data is not available. We evaluate prediction accuracy when we fine-tune the model using data from different seasons. We define the four fine-tuning periods shown in Table 2. We fine-tune

Table 2
Definitions of fine-tuning data periods.

Symbol	Period	Days
P1	2021.02.21–2021.04.30	32
P2	2021.05.01–2021.07.04	32
P3	2021.07.24–2021.10.26	30
P4	2021.10.27–2021.12.02	30

the model using only the data from each period.

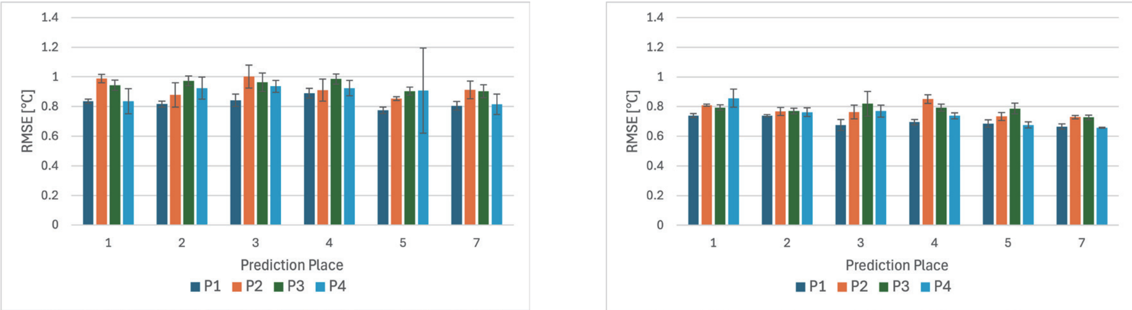
Figure 12 shows the results for winter prediction. Overall, the proposed method yields a lower *RMSE* than the comparison method. This result indicates that transfer learning improves prediction accuracy. However, compared with the case of using data from P1, which is the period immediately after winter, *RMSE* tends to increase when we use data from more temporally distant periods such as P2, P3, and P4. The p-values of the *t*-test indicate a significant difference at the 1% significance level in all cases of P1, while those for the other cases are relatively lower, where places 2 and 4 with P2, place 3 with P3, places 1, 5, and 7 with P4 do not have sufficient significance. This result suggests that transfer learning remains beneficial even when the fine-tuning data come from different seasons. At the same time, for winter prediction, fine-tuning with data closer to the evaluation period appears to be more effective.

Figure 13 shows the results for summer prediction. In the comparison method, *RMSE* becomes substantially worse at some sites when we train with data from P1 or P4, which correspond to seasons different from the evaluation period. In contrast, the proposed method does not show such large collapse when using data from P1 to P4. It maintains approximately stable accuracy across data of different fine-tuning periods, although the *t*-test results are weaker than the winter case, e.g., place 1 with P1, places 1 and 3–5 with P2, places 1 and 4 with P3, and places 2, 4, and 5 with P4 do not show significance at the 1% significance level. These results indicate that transfer learning suppresses performance degradation and enables robust estimation even when the season of the fine-tuning data does not match that of the evaluation period.

3.4.4 Ablation study of illuminance under different fine-tuning periods

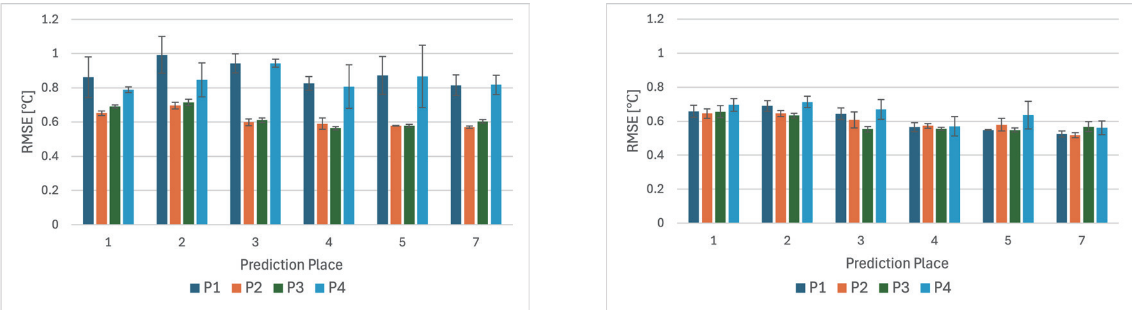
We conducted an ablation study to evaluate the effect of illuminance as a site-specific auxiliary feature. The experiment was conducted under the same setting as in Sect. 3.4.3, where the model was fine-tuned using one of the four periods P1–P4. We compared the transfer-learning model with illuminance and the same model without illuminance. The model architecture, pretraining data, fine-tuning periods, and evaluation periods were kept unchanged.

Figures 14 and 15 show the *RMSE* values for the winter and summer evaluation periods. Overall, the model with illuminance tended to achieve a lower *RMSE* than the model without illuminance in most combinations of target site and fine-tuning period. This result suggests that illuminance can provide useful site-specific information when the fine-tuning data are limited



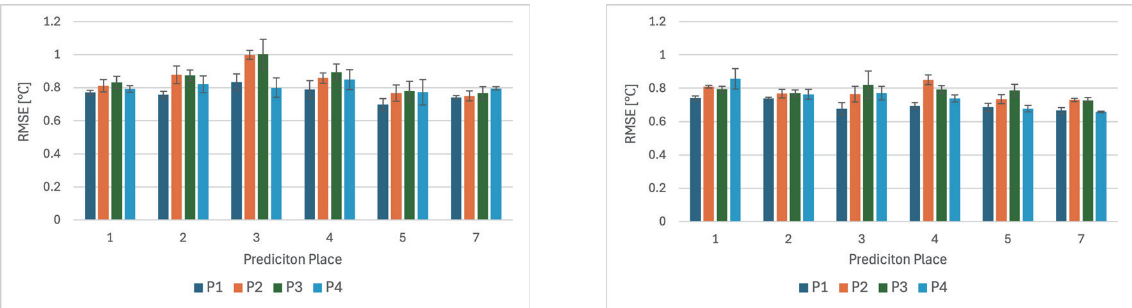
(a) (b)

Fig. 12. (Color online) *RMSE* values for winter prediction with different fine-tuning periods (P1–P4). (a) Comparison method (without transfer learning) and (b) proposed method (with transfer learning).



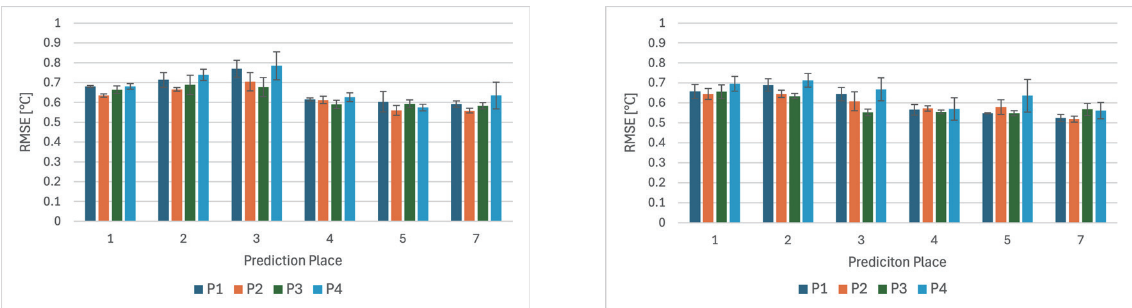
(a) (b)

Fig. 13. (Color online) *RMSE* values for summer prediction with different fine-tuning periods (P1–P4). (a) Comparison method (without transfer learning) and (b) proposed method (with transfer learning).



(a) (b)

Fig. 14. (Color online) Ablation comparison for winter prediction with different fine-tuning periods (P1–P4). (a) Comparison method (without illuminance) and (b) proposed method (with illuminance).



(a) (b)

Fig. 15. (Color online) Ablation comparison for summer prediction with different fine-tuning periods (P1–P4). (a) Comparison method (without illuminance) and (b) proposed method (with illuminance).

or seasonally different from the evaluation period.

However, the improvement was not statistically significant in all cases. The paired *t*-test showed significant differences at the 1% level in approximately half of the comparisons. Therefore, illuminance should not be regarded as a universally effective feature. Rather, its effectiveness depends on the target site, season, and fine-tuning period. This is reasonable because illuminance mainly represents solar radiation and shading, whereas local temperature is also affected by wind conditions, surface materials, vegetation, and surrounding building geometry.

3.4.5 Exploratory analysis of source-site selection for pretraining

In this subsection, we conduct an exploratory analysis of how the choice of source sites affects transfer learning. Thus far, we have used data from all six non-target sites for pretraining. We refer to this setting as all-site pretraining. We compare it with a setting that uses data from only one non-target site, which we refer to as single-site pretraining. The purpose of this analysis is not to determine the generally required amount of pretraining data, but to examine whether the source site used for pretraining affects prediction performance in the present dataset.

Here, pretraining means the stage in which we train the model using data from microclimate observation sites other than the target site. Fine-tuning means the stage in which we adapt the model using data from the target site. As in the previous subsection, we use the most recent 32 days of target-site data for fine-tuning.

Figure 16 shows the results. Each subfigure corresponds to one target site. The horizontal axis represents the site used for pretraining and the vertical axis represents *RMSE*. The bar chart includes both the winter and summer prediction results. For single-site pretraining, we evaluate all candidate source sites for each target site.

The results show that some single-site pretraining settings achieved *RMSE* values close to those obtained with all-site pretraining. In some cases, single-site pretraining even yielded a lower *RMSE* than all-site pretraining. These results suggest that, in the present dataset, adding more source-site data did not always lead to a lower *RMSE*. A possible reason is that source sites differ in their local environmental characteristics, and a source site that is more compatible with the target site may provide more useful transferable information than a heterogeneous mixture of all available sites.

Although local environments such as illuminance differ across sites, the gap between all-site pretraining and single-site pretraining remains small in most cases. This result indicates that the model can retain a certain level of prediction performance even when we reduce the pretraining data to a single site. This property is favorable for practical deployment under operational constraints.

At the same time, the site-specific feature introduced in this study is mostly limited to illuminance. Microclimate variation is affected not only by solar radiation but also by wind conditions, surface materials, surrounding vegetation, and building arrangement. If the model cannot sufficiently capture these intersite differences, transfer learning may even degrade prediction accuracy. Therefore, an important future issue is the analysis of site compatibility and optimization of the set of source sites used for pretraining.

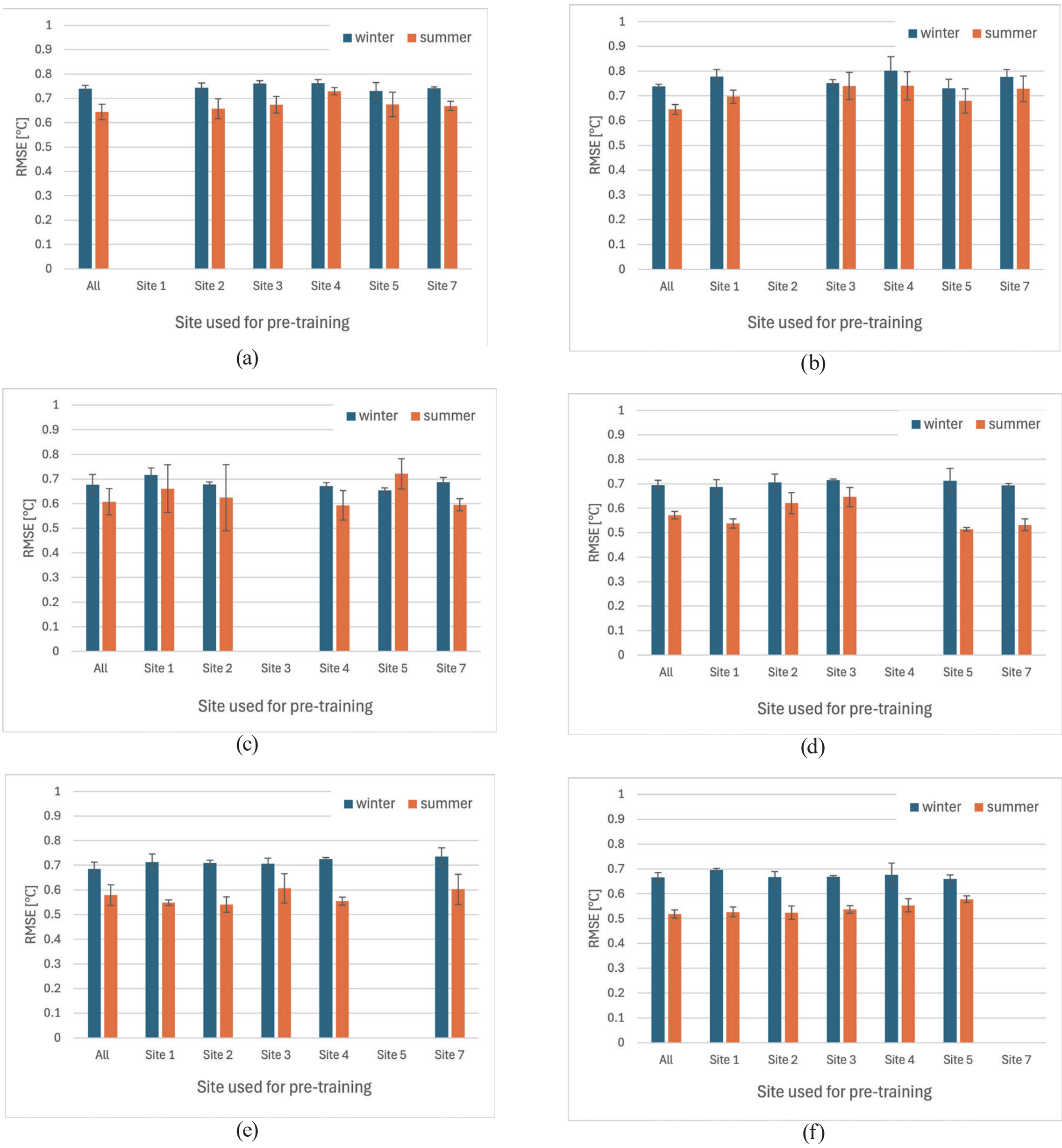


Fig. 16. (Color online) Prediction accuracy values for different source sites used in pretraining (winter and summer). (a) Site 1, (b) Site 2, (c) Site 3, (d) Site 4, (e) Site 5, and (f) Site 7.

4. Discussion

The results indicate that difference prediction is effective for the microclimate air-temperature prediction task considered in this study. A possible reason is that the nearby public reference station already captures the large-scale temporal variation in air temperature. Therefore, by predicting the temperature difference between the target site and the reference station, the model can focus on local deviations caused by the surrounding environment rather than learning the entire temperature sequence directly. This interpretation is consistent with the experimental results, where difference prediction improved accuracy compared with direct prediction under the same LSTM-based architecture.

The effect of transfer learning should be interpreted mainly in terms of robustness under limited-data conditions. Transfer learning did not always produce the lowest *RMSE* when sufficient target-site data were available. However, it was useful when the amount of fine-tuning data was limited or when the fine-tuning data were collected in a season different from the evaluation period. This suggests that pretraining on nearby microclimate sites can provide temporal representations that help stabilize prediction when target-site observations are insufficient. At the same time, the benefit of transfer learning depends on the relationship between the source and target sites, and it should not be assumed to improve accuracy in all conditions.

The ablation study of illuminance showed that illuminance can provide useful site-specific information in many cases, especially under limited or seasonally mismatched fine-tuning conditions. However, the improvement was not statistically significant in all comparisons. Therefore, illuminance should be regarded as a useful auxiliary feature, not as a universally effective feature. This is reasonable because illuminance mainly reflects solar radiation and shading, whereas microclimate air temperature is also affected by other local factors such as wind conditions, surface materials, vegetation, and surrounding building geometry. In addition, when illuminance is not observed at the public reference station, the representativeness of proxy illuminance should be carefully considered.

Another observation from the source-site selection experiment is that all-site pretraining did not always yield the lowest *RMSE*. Some single-site pretraining settings achieved performance comparable to, or better than, that of all-site pretraining. This suggests that source-site compatibility may affect transfer performance and that increasing the number of source sites does not necessarily improve prediction accuracy when the source sites have heterogeneous characteristics. However, this result should be interpreted only as an exploratory finding within the present campus-scale dataset. It does not imply that one source site is generally sufficient for pretraining.

This study has several limitations. The experiments were conducted using data from one campus environment with a limited number of target sites and observation days. Therefore, the findings should be validated using datasets from different urban forms, land-cover conditions, and climatic regions. In addition, this study used a fixed LSTM-based architecture to isolate the effects of difference prediction and transfer learning, and did not aim to conduct an exhaustive comparison of forecasting architectures. Future work should examine whether the same findings hold for other models and should incorporate additional site-specific features such as wind exposure, sky-view factor, vegetation, surface materials, and building geometry. A systematic method for selecting suitable source sites for pretraining is also an important future direction.

5. Conclusion

This study evaluated an LSTM-based framework for microclimate air-temperature prediction under limited observational resources. The framework combines difference prediction from a nearby public reference station with transfer learning from neighboring microclimate sites. The objective was to examine whether these strategies improve prediction accuracy and robustness in a data-limited microclimate prediction setting.

The results show that difference prediction improved accuracy compared with direct prediction under the same model architecture. Transfer learning was useful mainly when target-site data were limited or seasonally different from the evaluation period, although it did not always improve the best accuracy. The results also suggest that illuminance and source-site selection affected transfer performance, but these effects were not uniform across all sites and conditions.

These findings suggest that the proposed framework can support air-temperature prediction at newly installed or data-scarce microclimate sites under conditions similar to those examined in this study. Future work should validate the framework using more diverse datasets, incorporate additional site-specific features, and develop a systematic method for selecting suitable source sites for pretraining.

Acknowledgments

This work was supported by the Takahashi Industrial and Economic Research Foundation.

References

1 S. A. Zaki, S. W. Syahidah, M. F. Shahidan, M. I. Ahmad, F. Yakub, M. Z. Hassan, and M. Y. Md Daud: Sustainability **12** (2020) 5741. https://doi.org/10.3390/su12145741
2 Intergovernmental Panel on Climate Change (IPCC): https://openresearch-repository.anu.edu.au/bitstreams/95eb5ece-a734-44d7-86b3-bec919268af8/download (accessed March 2026).
3 World Meteorological Organization (WMO): https://wmo.int/guide-instruments-and-methods-of-observation-wmo-no-8-0 (accessed March, 2026).
4 F. Johannsen, P. M. M. Soares, and G. S. Langendijk: Urban Climate **56** (2024) 102039. https://doi.org/10.1016/j.uclim.2024.102039
5 Y. Yu, P. Li, D. Huang, and A. Sharma: Urban Climate **56** (2024) 102003. https://doi.org/10.1016/j.uclim.2024.102003
6 F. Chajaei and H. Bagheri: Urban Climate **57** (2024) 102102. https://doi.org/10.1016/j.uclim.2024.102102
7 P. Kumar, R. Chandra, C. Bansal, S. Kalyanaraman, T. Ganu, and M. Grant: Proc. 27th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (ACM 2021) 3128–3138. https://doi.org/10.1145/3447548.3467173
8 J. Han, A. Chong, J. Lim, S. Ramasamy, N. H. Wong, and F. Biljecki: Build. Environ. **253** (2024) 111358. https://doi.org/10.1016/j.buildenv.2024.111358
9 S. Yang, D. Zhan, T. Stathopoulos, J. Zou, C. Shu, and L. L. Wang: Energy Build. **314** (2024) 114283. https://doi.org/10.1016/j.enbuild.2024.114283
10 P. Shen: Energy Build. **343** (2025) 115923. https://doi.org/10.1016/j.enbuild.2025.115923
11 Q. Li, W. Wang, Z. Yu, and J. Chen: Build. Environ. **234** (2023) 110186. https://doi.org/10.1016/j.buildenv.2023.110186

12 I. Deznabi, P. Kumar, and M. Fiterau: arXiv:2401.02665 (2024).
13 Japan Meteorological Agency: https://www.jma.go.jp/jma/en/Activities/amedas/amedas.html (accessed March 2026).
14 S. Hochreiter and J. Schmidhuber: Neural Comput. **9** (1997) 1735. https://doi.org/10.1162/neco.1997.9.8.1735

About the Authors

Kiyoto Kimura received his B.E. degree from the Faculty of Systems Engineering, Wakayama University in 2024, and M.E. degree from the Graduate School of Systems Engineering, Wakayama University in 2026. He is interested in deep learning and environmental sensing. He is currently with Hitachi Solutions Ltd. He is a member of the Information Processing Society of Japan. (ki-mura.kiyoto@g.wakayama-u.jp)

Takuya Yoshihiro received his B.E., M.I., and Ph.D. degrees from Kyoto University in 1998, 2000, and 2003, respectively. He was an assistant professor from 2003 to 2009 and an associate professor from 2009 to 2023 at Wakayama University, where he is now a professor. His current interests include graph theory, distributed algorithms, mathematical optimization, information networks, IoT, intelligent transportation systems, data economics, medical applications, disaster recovery, climate prediction, and so forth. He is a member of IEEE, ACM, IEICE, and IPSJ. (tac@wakayama-u.ac.jp)