

<https://helda.helsinki.fi>

Image-to-Image Training for Spatially Seamless Air Temperature Estimation with Satellite Images and Station Data

SU, Peifeng

2023-03-13

SU, P, Abera, T, Guan, Y & Pellikka, P 2023, ' Image-to-Image Training for Spatially Seamless Air Temperature Estimation with Satellite Images and Station Data ', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing , vol. 16 , pp. 3353 - 3363 . <https://doi.org/10.1109/JSTARS.2023.3256363>

<http://hdl.handle.net/10138/357483>

<https://doi.org/10.1109/JSTARS.2023.3256363>

cc_by

publishedVersion


Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Image-to-Image Training for Spatially Seamless Air Temperature Estimation With Satellite Images and Station Data

Peifeng Su , Temesgen Abera, Yanlong Guan, and Petri Pellikka

Abstract—Air temperature at approximately 2 m above the ground (T_a) is one of the most important environmental and biophysical parameters to study various earth surface processes. T_a measured from meteorological stations is inadequate to study its spatio-temporal patterns since the stations are unevenly and sparsely distributed. Satellite-derived land surface temperature (LST) provides global coverage, and is generally utilized to estimate T_a due to the close relationship between LST and T_a . However, LST products are sensitive to cloud contamination, resulting in missing values in LST and leading to the estimated T_a being spatially incomplete. To solve the missing data problem, we propose a deep learning method to estimate spatially seamless T_a from LST that contains missing values. Experimental results on 5-year data of mainland China illustrate that the image-to-image training strategy alleviates the missing data problem and fills the gaps in LST implicitly. Plus, the strong linear relationships between observed daily mean T_a (T_{mean}), daily minimum T_a (T_{min}), and daily maximum T_a (T_{max}) make the estimation of T_{mean} , T_{min} , and T_{max} simultaneously possible. For mainland China, the proposed method achieves results with R^2 of 0.962, 0.953, 0.944, mean absolute error (MAE) of 1.793 °C, 2.143 °C, and 2.125 °C, and root-mean-square error (RMSE) of 2.376 °C, 2.808 °C, and 2.823 °C for T_{mean} , T_{min} , and T_{max} , respectively. Our study provides a new paradigm for estimating spatially seamless ground-level parameters from satellite products. Code and more results are available at <https://github.com/cvvsu/LSTa>.

Index Terms—Air temperature, deep learning, image-to-image mapping, land surface temperature, MODIS aqua, remote sensing.

I. INTRODUCTION

SURFACE air temperature (T_a), generally measured at approximately 2 m above the ground, is one of the most

Manuscript received 30 November 2022; revised 31 January 2023 and 7 March 2023; accepted 8 March 2023. Date of publication 13 March 2023; date of current version 7 April 2023. The work was supported in part by the MegaSense research programme of the University of Helsinki, in part by the City of Helsinki Innovation Fund, Business Finland, Business Finland Project under Grant 6884/31/2018 in part by MegaSense Smart City, in part by the European Commission through the Urban Innovative Action Healthy Outdoor Premises for Everyone under Grant UIA03-240, in part by SMARTLAND project funded by Academy of Finland under Grant 318645, and in part by Guangdong Basic and Applied Basic Research Fund under Grant 2022A1515011070. (Corresponding author: Peifeng Su.)

Peifeng Su, Temesgen Abera, and Petri Pellikka are with the Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland (e-mail: peifeng.su@helsinki.fi; temesgen.abera@helsinki.fi; petri.pellikka@helsinki.fi).

Yanlong Guan is with the School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518005, China (e-mail: guanyanlong2016@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3256363

important parameters for a wide range of research fields such as climatology [1], [2], hydrology [3], human health [4], and climate change [5]. The patterns of spatio-temporal variations of T_a are critical to understand several near-surface environmental and biotic processes as most terrestrial life lives within the near-surface of the earth [1].

The collected T_a from the meteorological stations is sparsely and unevenly distributed in the spatial dimension, leading to inadequate or biased comprehension of the spatio-temporal patterns of T_a that are strongly determined by the surface properties varying in both space and time [6], [7].

Spatially seamless T_a can be generated from weather station data by leveraging spatial interpolation methods such as Kriging and inverse distance weighting interpolations, while the interpolation accuracy may not be satisfying if the density of the weather stations is low or the topography is complex [8].

Remotely sensed datasets, such as the moderate resolution imaging spectroradiometer (MODIS) land surface temperature (LST) derived from the thermal infrared channels, provide global coverage to estimate T_a [9], [10], [11]. However, due to cloud contamination, the LST products contain missing data, making estimation methods, such as temperature-vegetation index methods, energy balance methods, and statistical methods (traditional machine learning and deep learning methods), mostly suitable for clear-sky conditions [12], [13], [14], [15], [16], [17], [18].

Microwave remote sensing and climate reanalysis datasets can also be applied to estimate spatially seamless T_a since they are insensitive to clouds. Nevertheless, the spatial resolutions of these datasets are typically rather coarse [19].

Daily mean T_a (T_{mean}), daily minimum T_a (T_{min}), and daily maximum T_a (T_{max}) are three attributes of T_a . Previous studies usually estimate T_{mean} , T_{min} , and T_{max} separately, which may omit the relationships between T_{mean} , T_{min} , and T_{max} . Additionally, estimating T_{mean} , T_{min} , and T_{max} separately makes the estimation process complex. In practice, estimating T_{mean} , T_{min} , and T_{max} one at a time will make data preparation laborious and error-prone, especially when different weather conditions are also considered [9], [18].

Since clouds cover a large area of the earth's surface [19], [20], a method that can directly use LST products that contain missing values to predict T_a is necessary. Inspired by image completion [21] and the image-to-image translation studies [22], we adopt an image-to-image training strategy with the UNet

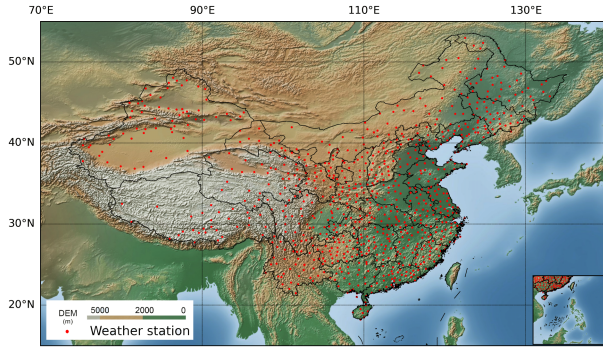


Fig. 1. Distribution of weather stations (red dots) applied in this article. The elevation in southwestern China is high (grey color). The density of stations is relatively high for eastern China while low for western China.

model to convert MODIS Aqua LST images containing missing values to spatially seamless T_a over mainland China during 2012–2016. The main findings and contributions of this study can be summarized as follows.

- 1) We find that T_{mean} , T_{min} , and T_{max} have strong linear relationships, illustrating the potential of estimating them simultaneously.
- 2) The image-to-image training strategy can grasp the spatial relationships in LST images and T_a images, which fills the gaps in LST images implicitly and obtains spatially seamless T_a .
- 3) Our study provides a new prototype to estimate spatially seamless ground-level parameters from satellite products, which is simple and straightforward. Experimental results validate that the MODIS LST products can also be used to estimate near-surface relative humidity (RH).

II. DATASETS AND METHODOLOGY

A. Study Area and Observed T_a

Mainland China is selected as the study area due to its vast and versatile land area. China covers approximately 9.6 million km^2 , between 3°N to 54°N latitudes, and 73°E to 136°E longitudes (see Fig. 1). China has a variety of climate types, such as tropical, subtropical, warm-temperate, temperate, cold-temperate, and highland climates. The elevation in China, ranging from -154 to 8848 m, is generally high in the west and low in the east.

In addition, China also has different land cover types, such as deserts, grasslands, croplands, forests, and urban areas [23]. The land cover types and topographical characteristics make the landscape in China versatile. Numbers of subcategories exist for each land cover type. For example, in China, forests include subtropical forests, boreal forests, subalpine forests, and so on.

There are 836 weather stations that offer the observed T_a (see Fig. 1), which is available from the China Meteorological Data Service Center (CMDC).¹

For the observed T_{mean} , T_{min} , and T_{max} , we discover that they statistically have strong linear relationships (see Fig. 2),

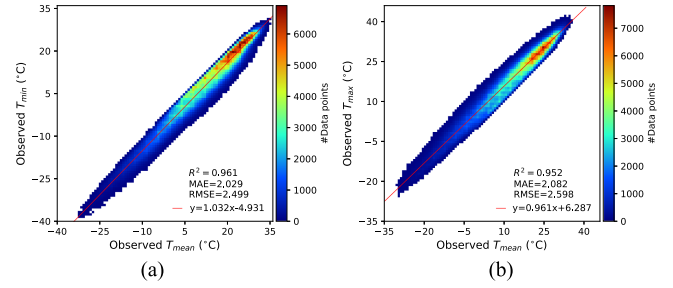


Fig. 2. Strong linear relationships between observed T_{mean} , T_{min} , and T_{max} over mainland China during 2012–2016. (a) Relationship between T_{mean} and T_{min} . (b) Relationship between T_{mean} and T_{max} .

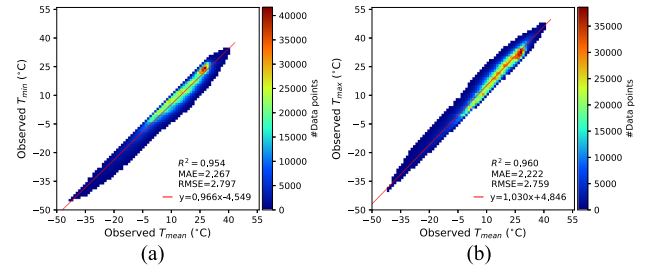


Fig. 3. Strong linear relationships for T_a measured worldwide in 2012.

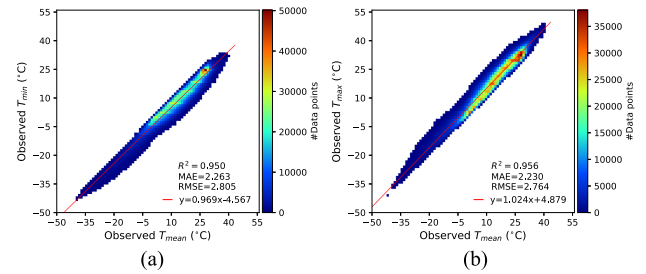


Fig. 4. Strong linear relationships for T_a measured worldwide in 2015.

illustrating the potential of estimating them simultaneously. In [18], a model is trained separately several times to estimate T_{mean} , T_{min} , and T_{max} . In [9], different models are trained to estimate T_{mean} under three different sky conditions, which is rather complex. Therefore, estimating T_{mean} , T_{min} , and T_{max} concurrently by leveraging only one model will be efficient. The strong linear relationships can not only be found for T_a measured in mainland China, but also worldwide.²

We display T_a collected from stations worldwide in 2012 and 2015 (see Fig. 3 and Fig. 4). To see the strong linear relationships in other years, please refer to our code. Thus the strong linear relationships seem to be a statistical property of T_a . Though the underlying mechanism is not clear, the linear property makes the estimation of T_{mean} , T_{min} , and T_{max} possible. The proof is given below.

²Relationships of measured T_{mean} , T_{min} , and T_{max} at the global scale: <https://github.com/cvvsu/LSTa/tree/main/datasets/images>

¹CMDC: <http://data.cma.cn/en>

Assuming that we are estimating T_{mean} y utilizing the satellite products x . That is, we find a model f

$$\hat{y} = f(x) \quad (1)$$

where y means the observed T_{mean} and \hat{y} means the estimated T_{mean} . We use coefficient of determination (R^2), mean absolute error (MAE), and root-mean-square error (RMSE) to evaluate the model f , and then

$$R_f^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$\text{MAE}_f = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (3)$$

$$\text{RMSE}_f = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

where \bar{y} is the mean of y .

Taking T_{min} (or T_{max}) z as an example, since z and y have a strong linear relationship, we have a linear model h , and

$$\hat{z} = h(y) = ay + b \quad (5)$$

where a and b are the parameters in the linear model h , and $a > 0$. Then

$$R_h^2 = 1 - \frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} = 1 - \frac{\sum_{i=1}^n (z_i - ay_i - b)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad (6)$$

$$\text{MAE}_h = \frac{\sum_{i=1}^n |z_i - \hat{z}_i|}{n} = \frac{\sum_{i=1}^n |z_i - ay_i - b|}{n} \quad (7)$$

$$\text{RMSE}_h = \sqrt{\frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (z_i - ay_i - b)^2}{n}} \quad (8)$$

To illustrate that it is possible to estimate T_{min} $\hat{z}p$ according to the strong linear relationships between z and y , we can suppose that

$$\hat{z}p = a\hat{y} + b \quad (9)$$

then

$$\begin{aligned} \text{MAE}_p &= \frac{\sum_{i=1}^n |z_i - \hat{z}p_i|}{n} \\ &= \frac{\sum_{i=1}^n |z_i - a\hat{y}_i - b|}{n} \\ &= \frac{\sum_{i=1}^n |z_i - ay_i - b + ay_i - a\hat{y}_i|}{n} \\ &\leq \frac{\sum_{i=1}^n |z_i - ay_i - b| + \sum_{i=1}^n |ay_i - a\hat{y}_i|}{n} \\ &= \text{MAE}_h + a\text{MAE}_f \end{aligned} \quad (10)$$

and

$$\begin{aligned} R_p^2 &= 1 - \frac{\sum_{i=1}^n (z_i - \hat{z}p_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \\ &= 1 - \frac{\sum_{i=1}^n (z_i - a\hat{y}_i - b)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \end{aligned}$$

$$= 1 - \frac{\sum_{i=1}^n (z_i - ay_i - b + ay_i - a\hat{y}_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2}. \quad (11)$$

According to Minkowski inequality [24]

$$\begin{aligned} &\sum_{i=1}^n (z_i - ay_i - b + ay_i - a\hat{y}_i)^2 \\ &\leq \left(\sqrt{\sum_{i=1}^n (z_i - ay_i - b)^2} + \sqrt{\sum_{i=1}^n (ay_i - a\hat{y}_i)^2} \right)^2 \\ &= \sum_{i=1}^n (z_i - ay_i - b)^2 + \sum_{i=1}^n (ay_i - a\hat{y}_i)^2 \end{aligned} \quad (12)$$

$$+ 2\sqrt{\sum_{i=1}^n (z_i - ay_i - b)^2 \sum_{i=1}^n (ay_i - a\hat{y}_i)^2}. \quad (13)$$

As a result

$$\begin{aligned} R_p^2 &= 1 - \frac{\sum_{i=1}^n (z_i - ay_i - b + ay_i - a\hat{y}_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \\ &\geq 1 - \frac{\sum_{i=1}^n (z_i - ay_i - b)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} - \frac{\sum_{i=1}^n (ay_i - a\hat{y}_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2} \\ &\quad - \frac{2\sqrt{\sum_{i=1}^n (z_i - ay_i - b)^2 \sum_{i=1}^n (ay_i - a\hat{y}_i)^2}}{\sum_{i=1}^n (z_i - \bar{z})^2} \\ &= 1 - (1 - R_h^2) - a^2(1 - R_f^2) - 2\sqrt{(1 - R_h^2)a^2(1 - R_f^2)} \\ &= R_h^2 + a^2R_f^2 - a^2 - 2a\sqrt{(1 - R_h^2)(1 - R_f^2)}. \end{aligned} \quad (14)$$

Similarly,

$$\begin{aligned} \text{RMSE}_p &= \sqrt{\frac{\sum_{i=1}^n (z_i - \hat{z}p_i)^2}{n}} \\ &= \sqrt{\frac{\sum_{i=1}^n (z_i - ay_i - b + ay_i - a\hat{y}_i)^2}{n}} \\ &\leq \sqrt{\frac{\sum_{i=1}^n (z_i - ay_i - b)^2}{n}} + a\sqrt{\frac{\sum_{i=1}^n (ay_i - a\hat{y}_i)^2}{n}} \\ &= \text{RMSE}_h + a\text{RMSE}_f. \end{aligned} \quad (15)$$

The above proof illustrates that within some error range (R^2 , MAE, and RMSE), we can estimate T_{mean} , T_{min} , and T_{max} simultaneously just by applying the strong linear relationships.

B. Remotely Sensed Data

We use LST data from the MYD11A1 version 6.1 daily product from MODIS Aqua satellite as the overpass times of MODIS Aqua are quite near the times that T_{min} and T_{max} occur [25].

The daytime LST (LSTD) and nighttime LST (LSTN) products are downloaded from Google Earth Engine [26] with a Python package *geemap* [27]. Since we try to estimate T_{mean} , T_{min} , and T_{max} simultaneously, both the Aqua LSTD and LSTN products are used. The daily LSTD and LSTN are stacked to obtain LST images. Suppose that the size of the LSTD image is $m \times n$ pixels, then the size of an LST image in this article is $m \times n \times 2$ pixels. Because the study area is quite large and

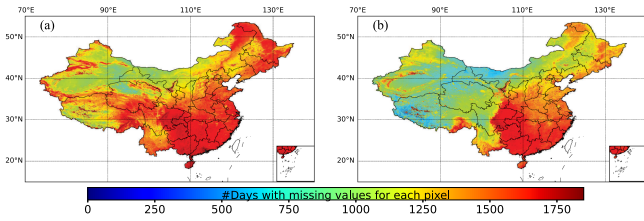


Fig. 5. Number of days for LSTD and LSTN that are with missing values. For each pixel, (a) and (b) show the number of days that have no valid data for LSTD and LSTN, respectively. There are 1827 days in 2012–2016.

if 1 km spatial resolution were applied, the image size of an LST image would be $3492 \times 6862 \times 2$ pixels. Thus, we reproject the LST products to the World Geodetic System 1984 (EPSG: 4326) and resample LST images to 5 km spatial resolution ($789 \times 1373 \times 2$ pixels).

Only the good-quality pixels whose average LST error is less than or equal to 1 K are kept, and other pixels are seen as pixels with missing values. Since southeastern China is with high humid, most of the days for LSTD and LSTN have no valid value (see Fig. 5). In [9] and [14], the missing values in LST products are replaced using temporal gap-filling methods. But for the study area utilized in this article, some missing data cannot be replaced since valid temporal adjacent LST values are not available (see Fig. 5).

Contrary to [9] and [14], we replace the missing values in LST images with a constant value rather than a temporal nearby value. The filled value in this study has *no physical meaning* and is only used for numeric computation. Since this investigation is not an LST gap-filling study, the filled value has no effect on the real LST products. The effects on estimation results caused by the filled value will be discussed in Section III-A.

C. Previous Statistical Methods

There are many other auxiliary datasets, such as normalized difference vegetation index (NDVI), that can be used together with LST for T_a estimation. For simplicity, we only use LST to illustrate the estimation steps of previous statistical methods.

To estimate spatially seamless T_a , no matter which machine or deep learning model is applied, there should be a method to impute the missing values in LST in advance [9], [14]. After imputing the missing values, then the data pairs between LST and T_a were established (green pixels in LST and T_{mean} images as shown in Fig. 6). A model was trained based on the data pairs and then applied to other pixels with valid values (blue pixels in the LST image, Fig. 6). However, as shown in Figs. 5 and 6, there may be pixels that cannot be filled by nearby clear-sky values when there is an area where most pixels are with missing values. Furthermore, if a model was trained on the data pairs, it may lose the spatial relationships between pixels in LST images and the spatial relationships in T_a images.

Therefore, previous methods have to use land cover or elevation information to explicitly illustrate the spatial relationships to the model. In fact, the information about NDVI, land cover, and elevation has already been implicitly encoded into the LST

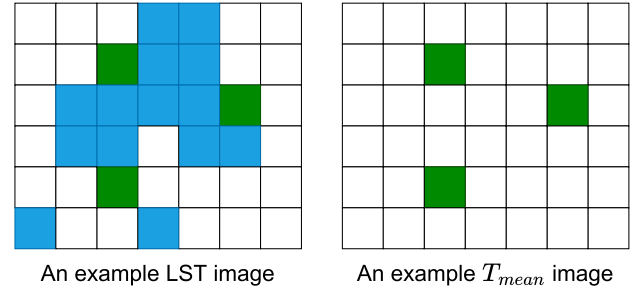


Fig. 6. Example LST and T_{mean} images to illustrate data pairs, clear-sky LST values, and missing values. The green pixels are data pairs between LST and T_{mean} images. The blue pixels are with clear-sky LST values, and the white pixels are with missing values.

products. That is, different land types or elevation leads to different LST values.

D. Daily T_a Images

The T_a (T_{mean} , T_{min} , T_{max}) measured in meteorological stations are data samples. Since T_a are not independent in space, we convert the observed T_a samples to a T_a image for each day (see Fig. 7) to keep the spatial relationships between the observed T_a samples. The detailed steps are as follows.

- 1) An image with the size of $m \times n \times 3$ is created (stacked T_{mean} , T_{min} , and T_{max} images, respectively), and all pixels are given a constant value.
- 2) The values of those pixels whose geographical locations coincided with the location of meteorological stations are reallocated by the observed T_{mean} , T_{min} , and T_{max} .

The constant value in T_a images will not affect the estimation results as the specific value will not participate in the loss calculation (illustrated in Section II-G).

E. Image Completion

Matrix completion is a method to predict the missing entries in a matrix by assuming that the matrix is highly sparse [28], [29], [30], and it has been widely used in many fields, such as computer vision and recommendation systems [31].

The sparsity property is also known as a prior for images [32], [33], [34]. According to [35], convolutional neural networks (CNNs) are popular tools for image generation and restoration as they can learn realistic image priors. For instance, CNNs are able to restore an image with a complex degradation [35], such as denoising, super-resolution, and inpainting.

Both studies on image inpainting and matrix completion demonstrate that there are priors in matrices and images, and according to these priors, it is possible to recover the missing entries in matrices and images. That is

$$x = f_0(x_0) \quad (16)$$

where x is a matrix or an image, x_0 is x with missing entries, and f_0 is a matrix or image completion method.

Though LST and RGB images have different imaging mechanisms, they share some common priors, such as sparsity and locality. In this case, we can use CNNs to complete the missing

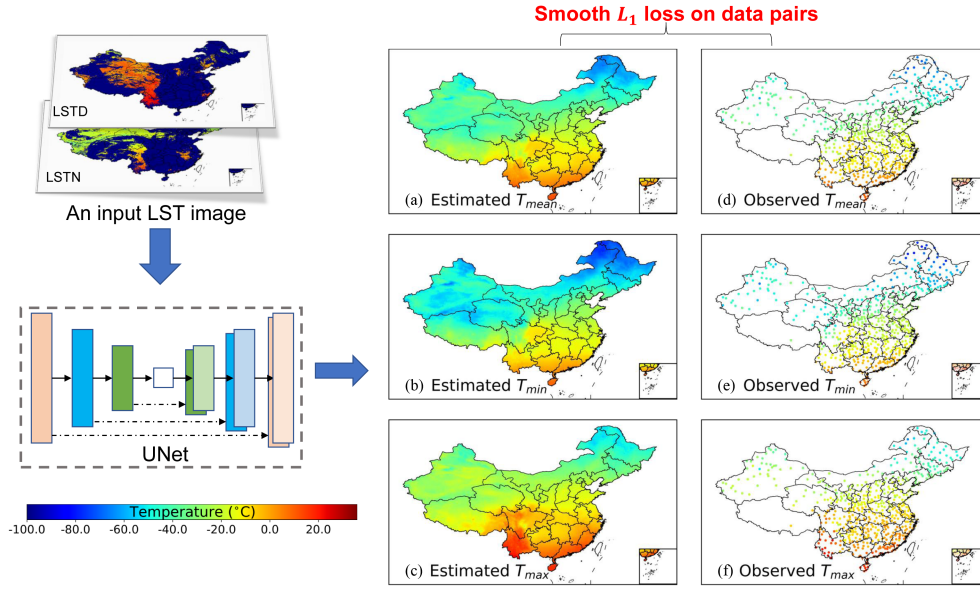


Fig. 7. Illustration of the image-to-image training strategy with UNet. Given an input LST image, the UNet model will output the estimated (a) T_{mean} , (b) T_{min} , and (c) T_{max} simultaneously. During the training phase (585 stations), the loss is calculated between the data pairs, and thus the filled value in T_a images (d, e, and f) has no effect on estimation results.

entries in LST images and then use the gap-filled results to estimate spatially seamless T_a .

Since we only focus on T_a estimation, a composite function f can be applied to estimate T_a from LST images with missing values directly

$$\hat{T}_a = f(x) = f_1(f_0(x_0)) \quad (17)$$

where $x_0 \in \mathcal{R}^{m \times n \times 2}$ in this study is an input LST image, f_1 is a mapping function that represents the relationships between spatially seamless LST images and T_a images, and $\hat{T}_a \in \mathcal{R}^{m \times n \times 3}$ is an estimated T_a image. We need to make sure that the estimated T_a images are as close as possible to the observed T_a for paired pixels. Thus, the inpainting processing for input LST images is implicitly performed. It should be noted that, unlike image and matrix completion methods, the composite function f should also contain the priors in observed T_a .

F. UNet and Image-to-Image Training Strategy

As the relationship between LST and T_a is complex and nonlinear, advanced machine learning or deep learning models can achieve better estimation results than linear models [7], [14], [17].

Deep learning techniques have powered various study areas, such as remote sensing, computational physics, and environment [36], [37], [38], [39], [40], [41]. UNet [42] has been widely applied to remote sensing and medical image processing according to its capability on pixel-level tasks such as semantic segmentation [43]. UNet is a specific encoder-decoder architecture with skip connections, and through the skip connections between mirrored layers, both the low-level and semantic features are available for the final estimation [22]. Therefore, we adopt the UNet model to perform the image-to-image training for spatially

seamless T_a estimation (see Fig. 7). In other words, UNet is utilized to learn the composite mapping function f .

For each day in 2012–2016 (1827 days in total), the input is an LST image with the size of $789 \times 1373 \times 2$ pixels and the output is a T_a image with the size of $789 \times 1373 \times 3$ pixels.

G. Training Details

We randomly split the 836 stations into training, validation, and test sets with a ratio of 70% (585): 10% (83): 20% (168). To avoid overfitting, during the training phase, we pad the input images and then randomly crop the desired size out (789×1373). As a result, the UNet model only “see” a portion of the whole input image during the training phase. Horizontal random flip transformation is also applied during training.

There are 200 epochs and the batch size is 8. The initial learning rate is 1.6×10^{-4} , and the cosine annealing schedule for the learning rate is utilized [44]. Please refer to the code³ to see more training details and results.

Only the data pairs are utilized for loss computation. The smooth L_1 loss l_{smooth} [45] is used as the loss function

$$l_{smooth} = \begin{cases} 0.5 * (y - \hat{y})^2, & |y - \hat{y}| < 1 \\ |y - \hat{y}| - 0.5, & otherwise \end{cases} \quad (18)$$

where y is the observed value and \hat{y} is the estimated value. Thus, the filled value in T_a images have no effect on estimation results (see Fig. 7).

R^2 , MAE, and RMSE are applied to evaluate the estimation results. Only the test results are reported if not specified.

³Code: <https://github.com/cvvsu/LSTa>

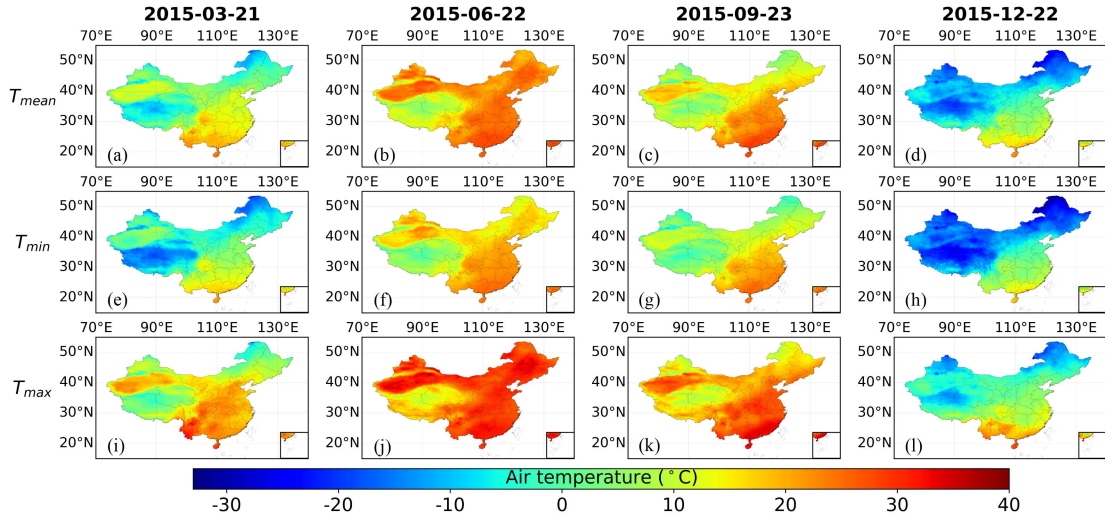


Fig. 8. Spatially seamless estimation results on four specific days in 2015 for T_{mean} , T_{min} , and T_{max} over mainland China, respectively. Estimation results for T_{mean} , T_{min} , and T_{max} are shown from the top row to the bottom row. Estimation results for the March equinox (March 21, 2015), the Summer solstice (June 22, 2015), the September equinox (September 23, 2015), and the Winter solstice (December 12, 2015) are shown from the left column to the right column. Note that we do not provide the elevation and land cover information for the proposed method, but the proposed method automatically captures these pieces of information. Zoom in to see the differences of T_{max} for Summer solstice.

TABLE I

ESTIMATION RESULTS ON THE VALIDATION SET BASED ON DIFFERENT FILLED VALUES IN LST IMAGES

Filled value (°C)	T_{mean}			T_{min}			T_{max}		
	R^2	MAE (°C)	RMSE (°C)	R^2	MAE (°C)	RMSE (°C)	R^2	MAE (°C)	RMSE (°C)
-100	0.962	1.731	2.311	0.952	2.056	2.686	0.944	2.070	2.770
-200	0.960	1.773	2.360	0.951	2.085	2.714	0.941	2.142	2.858
-500	0.957	1.836	2.436	0.949	2.133	2.779	0.938	2.193	2.917

TABLE II

ESTIMATION RESULTS FOR T_{mean} , T_{min} , AND T_{max} BASED ON INPUT LST IMAGES

	R^2	MAE (°C)	RMSE (°C)
T_{mean}	0.962	1.793	2.376
T_{min}	0.953	2.143	2.808
T_{max}	0.944	2.125	2.823

III. RESULTS AND DISCUSSION

A. Effects Caused by Filled Values

The filled value in LST images is a hyperparameter, and the validation set is utilized to determine the value. To make the filled value distinguishable from the valid value ranges in LST images, we use -100 , -200 , and -500 °C to fill the gaps. According to Table I, the filled values do not have a significant effect on the estimation results, in terms of R^2 , MAE, and RMSE values. We use -100 °C to fill all the missing data in LST images since it achieves the best estimation results.

B. Overall Estimation Results

As shown in Table II, the R^2 values are 0.962, 0.953, 0.944, and the MAE values are 1.793 °C, 2.143 °C, 2.125 °C, and the RMSE values are 2.376 °C, 2.808 °C, and 2.823 °C for T_{mean} , T_{min} , and T_{max} , respectively. Though T_{mean} , T_{min} , and T_{max} are estimated concurrently, the estimation results of T_{mean} is the better than that of T_{min} and T_{max} .

We select four special days in 2015 to visualize the estimation results (see Fig. 8) and videos that display the daily changes of T_{mean} , T_{min} , and T_{max} during 2012–2016 is available online.⁴ Note that the Summer solstice is the hottest day in a year, and the figure cannot display distinguish difference for T_{max} in most regions (please zoom in to see the differences in different areas, Fig. 8). The estimated results display elevation information and clear seasonal changes and daily changes of T_a (please see Fig. 8 and online videos).

Fig. 8 and the online videos verify that the proposed method can implicitly fill the missing data in LST images and estimate spatially seamless T_{mean} , T_{min} , and T_{max} , simultaneously. The filling process is similar to image completion studies [21], while we do not intend to fill the gaps in LST images but T_a images. Though we only apply the LST images as the input and no other auxiliary datasets, such as normalized difference vegetation index and land cover, are used, the estimated T_a displays the spatial heterogeneity in the study area. Southwestern China has a high elevation (the Qinghai-Tibet Plateau, Fig. 1), and the estimated T_a in this area shows relatively lower T_a compared with its nearby regions (see Fig. 8), indicating that the proposed method also captures the elevation information automatically. The possible reason is that the image-to-image training strategy grasps the spatial relationships in LST images and T_a images.

⁴Videos for estimated T_{mean} , T_{min} , and T_{max} : <https://github.com/cvvsu/LSTa/releases/tag/videos>

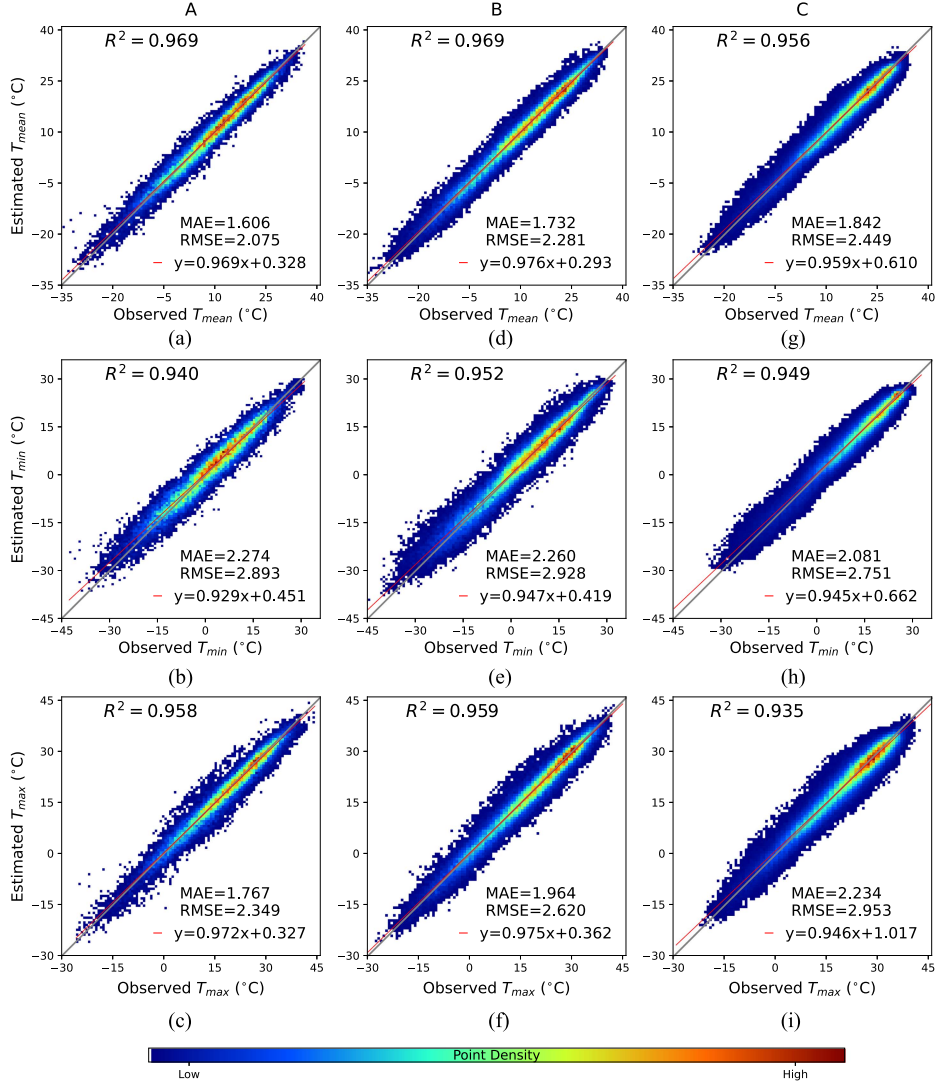


Fig. 9. Estimation results on pixels under different sky conditions. (a)–(c) Estimation results on pixels with clear-sky values in LST (part A). (d)–(f) Estimation results on pixels with one missing value in LSTD or LSTN (part B). (g)–(i) Estimation results on pixels with two missing values (part C). Estimation results for T_{mean} , T_{min} , and T_{max} are shown from the top to the bottom row. The gray line in each subplot represents $y = x$.

C. Effects of Missing Values

To show the effects caused by the missing values, we split T_a into three parts: the paired pixels in LST are recorded under clear-sky conditions (part A); the paired pixels in LSTD or LSTN are with missing values (part B); the paired pixels in LST are with missing values (part C). Note that LST images have two channels (see Fig. 7).

According to Fig. 9, the missing values in LST do not affect the estimation accuracy significantly. Specifically, the R^2 ranges from 0.956 to 0.969 for T_{mean} , 0.940 to 0.949 for T_{min} , and 0.935 to 0.959 for T_{max} , respectively. The MAE values fluctuate from 1.606 °C to 1.842 °C, 2.081 °C to 2.274 °C, and 1.767 °C to 2.234 °C for T_{mean} , T_{min} , and T_{max} , respectively. Similarly, RMSE values range from 2.075 °C to 2.449 °C for T_{mean} , 2.751–°C to 2.928 °C for T_{min} , and 2.349 °C to 2.953 °C for T_{max} , respectively. The above results validate that the proposed method alleviates the problem caused by missing values in LST.

D. Estimation Results Analysis

To analyze the estimate results, MAE values over space (see Fig. 10), time (see Fig. 11), and elevation (see Fig. 11) are calculated.

The MAE values of most stations are less than 3.000 °C (see Fig. 10), illustrating that the estimation results are stable in most regions. The stations with larger MAE values (≥ 3.000 °C) are mostly located in western China (see Fig. 10) where the density of stations is low, and stations with smaller MAE values (1.000 °C–2.000 °C) are located in eastern China where the density of stations is high. However, since the topography is much more complex in western China (see Fig. 1), we cannot state that the density of stations is the main reason accounting for the estimation accuracy with 100% confidence. For instance, in a flat region, the representativeness of a station should be better than the station in a region whose topography is complex. Thus, the representativeness of a station is not only determined

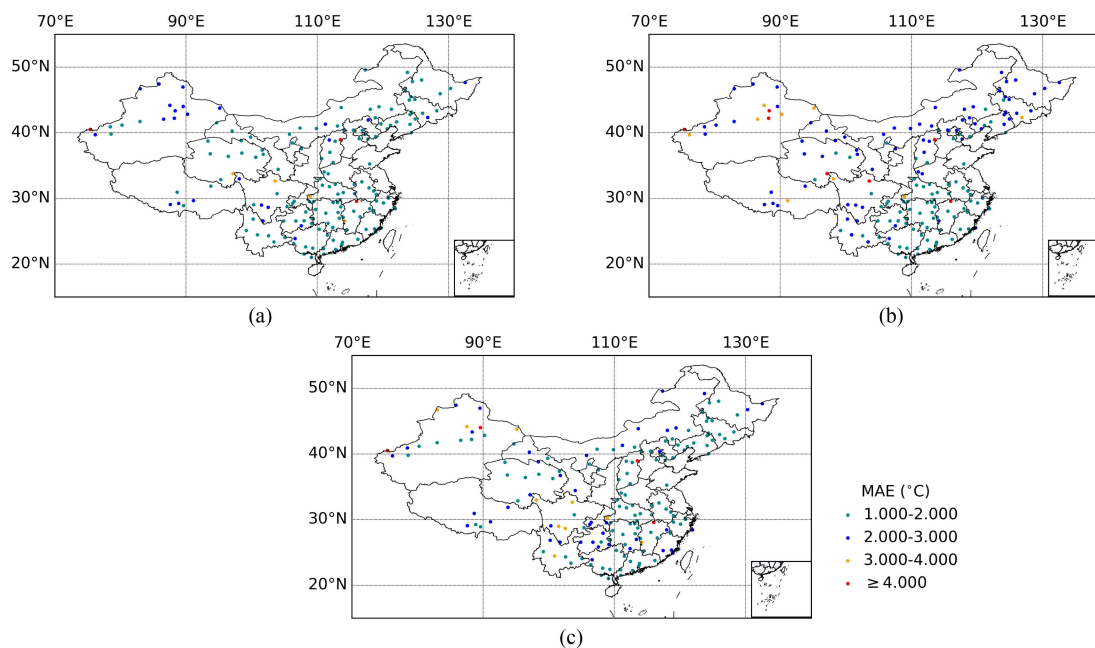


Fig. 10. MAE values for estimated T_a over space (test set: 168 stations).

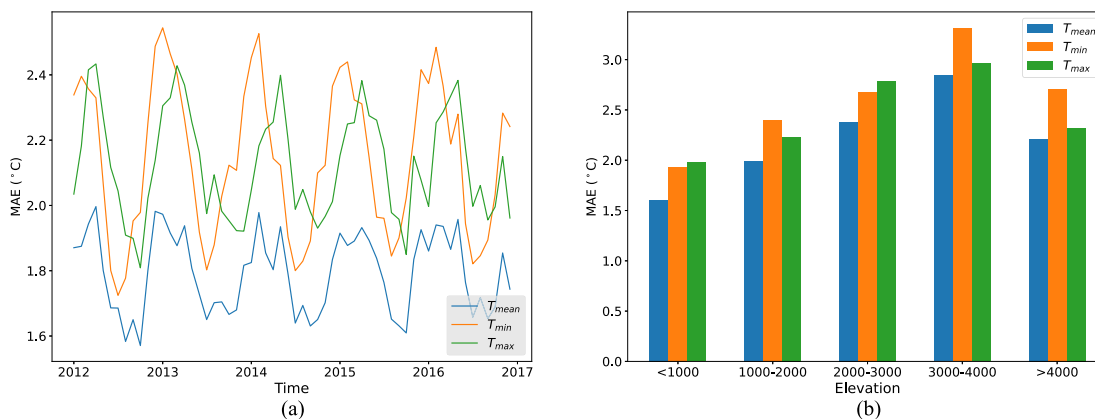


Fig. 11. MAE values over time and elevation for estimation results (test set: 168 stations).

by the number of stations nearby, but also by the topographic complexity.

The MAE values are similar between different years, following the same pattern: larger MAE values in winter and lower MAE values in summer (see Fig. 11). The reason may be that in winter the variation of T_a is higher compared with the variation in summer, making T_a collected in winter more difficult to be estimated. The seasonal pattern that T_a changes more significantly in winter than summer has also been reported in the United States.⁵ Another possible reason is that the relationship between LST and T_a is strong in late summer and fall, and weak in winter and early spring [46].

The MAE values rise first and then fall with the increase of elevation (see Fig. 11). However, the estimation accuracy is relatively stable at different elevations as the MAE values are generally less than 3.000 °C.

E. Estimation of Other Parameters

Since relative humidity (RH, whose unit is %) is highly related to T_a , we use the LST images to estimate RH by leveraging UNet and the image-to-image training method. On the test set, the estimation results are with R^2 of 0.713, MAE of 6.725%, and RMSE of 8.713% (see Fig. 12). To the best of our knowledge, we are the first ones who try to estimate spatially seamless RH from MODIS LST products.⁶ The estimation results of RH show the

⁵[Online]. Available: <https://www.epa.gov/climate-indicators/climate-change-indicators-seasonal-temperature>

⁶A video for estimated RH: <https://github.com/cvvsu/LSTa/releases/tag/videos>

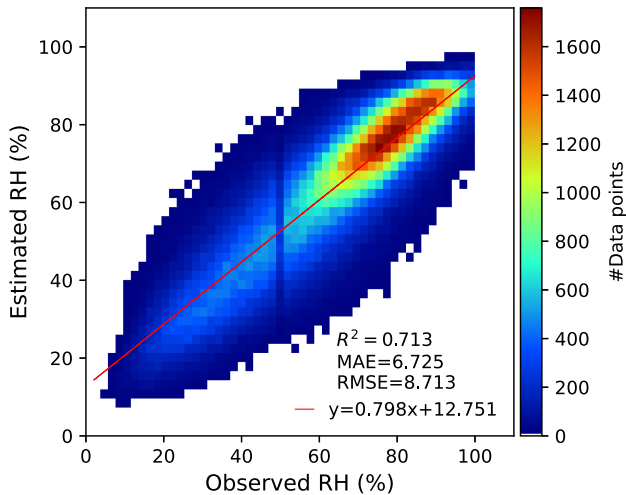


Fig. 12. Estimation results of RH based on LST images. % is the unit of RH.

potential of using other satellite products that contain missing data to estimate spatially seamless near-surface parameters. For instance, aerosol optical depth (AOD) products, which can be used to estimate $PM_{2.5}$ [47], [48], [49], are also sensitive to cloud contamination. By leveraging the proposed method, it should be possible to obtain spatially seamless $PM_{2.5}$ from the AOD products that contain missing data.

F. Comparison With Other Methods

In 2022, a method is proposed in [18] to estimate T_{mean} , T_{min} , and T_{max} by training a model several times. The RMSE values reported in [18] are in the range of 2.15 °C–3.20 °C for T_{max} , 1.68 °C–2.79 °C for T_{min} , and 0.86 °C–1.60 °C for T_{mean} for five different parts of mainland China. We use only one model and we do not use other auxiliary products, while the error level is similar for the proposed method and the method proposed in [18]. The climate reanalysis dataset ERA5 [50] has the RMSE values of 2.543 °C, 3.324 °C, and 3.393 °C for T_{mean} , T_{min} , and T_{max} , respectively. Compared with ERA5 and Fang’s results, our method is simple yet efficient. In addition, the spatial resolution for the proposed method is 5 km, while it is around 27.8 km for ERA5 and 10 km for Fang’s results.

In [7] and [9], 1 km spatial resolution T_{max} and T_{mean} are estimated, respectively. However, they directly split the data pairs to train models, and maybe independent stations are needed to verify the spatial estimation results of their methods. Plus, the estimation results in [7] are spatially incomplete, and in [9], there is no value for water-body regions.

Despite that our method is simple and straightforward, our method considers the spatial relationships between pixels in LST images and the spatial relationships between observed T_a . As a result, we do not have to explicitly provide other variables, such as land cover and elevation, to the model to obtain estimated T_a .

We only consider the worst case for estimation. For instance, we only use MODIS Aqua products, and according to [14], the combination of MODIS Terra and Aqua products will obtain better estimation results. We use the LST error flag ≤ 1 K, while

we find that the 3 K error flag will provide better estimation results. This is not surprising since these pixels within the 3 K error flag are much better than the pixels filled with the constant value (-100 °C). In this case, we can say that our method would obtain better results if more LST products and ERA5 datasets were utilized as the input variables. However, to obtain the best numeric accuracy is not our purpose, and we try to *provide a new paradigm* for the estimation of T_a and other surface parameters.

Our method also works well on fine-scale satellite products, and please refer to our code to check the estimation results for Finland with a 1 km spatial resolution.

IV. CONCLUSION

In this article, we propose a deep learning method to estimate spatially seamless T_a (T_{mean} , T_{min} , and T_{max} simultaneously) with the presence of missing data in LST products. Encouraging estimation results are obtained with an image-to-image training strategy and the UNet model. The estimation results of RH illustrate that the proposed method provides a new paradigm for estimating near-surface parameters from satellite products that contain missing values.

In the future, the question of whether the topographic complexity or the density of weather stations contributes most to the estimation accuracy should be further investigated. Additionally, the underlying mechanisms of why T_{mean} , T_{min} , and T_{max} statistically have strong linear relationships need to be explored and the image-to-image training strategy calls for lightweight convolutional neural networks to deal with large-scale remote sensing images.

ACKNOWLEDGMENT

The authors would like to thank CSC-IT Center for Science, Finland, for computational resources and the editors and three anonymous reviewers for their constructive comments and suggestions.

REFERENCES

- [1] L. Prihodko and S. N. Goward, “Estimation of air temperature from remotely sensed surface observations,” *Remote Sens. Environ.*, vol. 60, no. 3, pp. 335–346, 1997.
- [2] I. Kloog et al., “Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data,” *Int. J. Climatol.*, vol. 37, no. 1, pp. 296–304, 2017.
- [3] O. Mohseni and H. Stefan, “Stream temperature/air temperature relationship: A physical interpretation,” *J. Hydrol.*, vol. 218, no. 3/4, pp. 128–141, 1999.
- [4] L. Shi, P. Liu, I. Kloog, M. Lee, A. Kosheleva, and J. Schwartz, “Estimating daily air temperature across the southeastern United States using high-resolution satellite data: A statistical modeling study,” *Environ. Res.*, vol. 146, pp. 51–58, 2016.
- [5] M. Gao, Z. Li, Z. Tan, H. Li, and J. Peng, “Use of Google Earth Engine to generate a 20-year 1 km \times 1 km monthly air temperature product over Yellow River Basin,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10079–10090, 2021.
- [6] A. Rosenfeld, M. Dorman, J. Schwartz, V. Novack, A. C. Just, and I. Kloog, “Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel,” *Environ. Res.*, vol. 159, pp. 297–312, 2017.

- [7] H. Shen, Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang, "Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data," *Remote Sens. Environ.*, vol. 240, 2020, Art. no. 111692.
- [8] Y. Shi, Z. Jiang, L. Dong, and S. Shen, "Statistical estimation of high-resolution surface air temperature from MODIS over the Yangtze River Delta China," *J. Meteorological Res.*, vol. 31, no. 2, pp. 448–454, 2017.
- [9] Y. Chen, S. Liang, H. Ma, B. Li, T. He, and Q. Wang, "An all-sky 1 km daily land surface air temperature product over mainland China for 2003–2019 from MODIS and ancillary data," *Earth Syst. Sci. Data*, vol. 13, no. 8, pp. 4241–4261, 2021.
- [10] F. Hong et al., "A simple yet robust framework to estimate accurate daily mean land surface temperature from thermal observations of tandem polar orbiters," *Remote Sens. Environ.*, vol. 264, 2021, Art. no. 112612.
- [11] A. Jia, H. Ma, S. Liang, and D. Wang, "Cloudy-sky land surface temperature from VIIRS and modis satellite data using a surface energy balance-based method," *Remote Sens. Environ.*, vol. 263, 2021, Art. no. 112566.
- [12] W. Zhu, A. Lü, and S. Jia, "Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products," *Remote Sens. Environ.*, vol. 130, pp. 62–73, 2013.
- [13] R. Zhang, Y. Rong, J. Tian, H. Su, Z.-L. Li, and S. Liu, "A remote sensing method for estimating surface air temperature and surface vapor pressure on a regional scale," *Remote Sens.*, vol. 7, no. 5, pp. 6005–6025, 2015.
- [14] Y. Rao et al., "Estimating daily average surface air temperature using satellite land surface temperature and top-of-atmosphere radiation products over the tibetan plateau," *Remote Sens. Environ.*, vol. 234, 2019, Art. no. 111462.
- [15] B. Zhou et al., "Estimation of hourly near surface air temperature across Israel using an ensemble model," *Remote Sens.*, vol. 12, no. 11, 2020, Art. no. 1741.
- [16] R. S. Dos Santos, "Estimating spatio-temporal air temperature in london (U.K.) using machine learning and earth observation satellite data," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 88, 2020, Art. no. 102066.
- [17] H. Zhang, W. Immerzeel, F. Zhang, R. J. De Kok, S. J. Gorrie, and M. Ye, "Creating 1-km long-term (1980–2014) daily average air temperatures over the Tibetan Plateau by integrating eight types of reanalysis and land data assimilation products downscaled with modis-estimated temperature lapse rates based on machine learning," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 97, 2021, Art. no. 102295.
- [18] S. Fang et al., "Dataset of daily near-surface air temperature in China from 1979 to 2018," *Earth Syst. Sci. Data*, vol. 14, no. 3, pp. 1413–1432, 2022.
- [19] H. Li, H. Liu, M. Duan, X. Deng, and S. Zhang, "Estimation of air temperature under cloudy conditions using satellite-based cloud products," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [20] B. Yao et al., "Can atmospheric reanalyses (CRA and ERA5) represent cloud spatiotemporal characteristics?," *Atmospheric Res.*, vol. 244, 2020, Art. no. 105091.
- [21] C. Zheng, T.-J. Cham, and J. Cai, "Pluralistic image completion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1438–1447.
- [22] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.
- [23] H. Li et al., "Using land long-term data records to map land cover changes in China over 1981–2010," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1372–1389, Apr. 2017.
- [24] E. Kreyszig, *Introductory Functional Analysis With Applications*, vol. 17. New York, NY, USA: Wiley, 1991.
- [25] R. Alkama and A. Cescatti, "Biophysical climate impacts of recent changes in global forest cover," *Science*, vol. 351, no. 6273, pp. 600–604, 2016.
- [26] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, 2017.
- [27] Q. Wu, "Geemap: A Python package for interactive mapping with google earth engine," *J. Open Source Softw.*, vol. 5, no. 51, 2020, Art. no. 2305.
- [28] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, pp. 717–772, 2008.
- [29] R. Bell, Y. Koren, Y. Research, and I. Volinsky, "The bellkor 2008 solution to the netflix prize," *Statist. Res. Dept. AT&T Res.*, vol. 1, no. 1, 2008.
- [30] L. Kassab, H. Adams, and D. Needell, "An adaptation for iterative structured matrix completion," in *Proc. 54th Asilomar Conf. Signals Syst. Comput.*, 2020, pp. 1451–1456.
- [31] X. Li, H. Zhang, and R. Zhang, "Matrix completion via non-convex relaxation and adaptive correlation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1981–1991, Feb. 2023.
- [32] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [33] I. Higgins et al., "Beta-vae: Learning basic visual concepts with a constrained variational framework," in *Proc. 5th Int. Conf. Learn. Representations*, 2017. [Online]. Available: <https://openreview.net/forum?id=Sy2fzU9gl>
- [34] C. Nash, J. Menick, S. Dieleman, and P. W. Battaglia, "Generating images with sparse representations," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 7958–7968. [Online]. Available: <http://proceedings.mlr.press/v139/nash21a.html>
- [35] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Deep image prior," *Int. J. Comput. Vis.*, vol. 128, no. 7, pp. 1867–1888, 2020. [Online]. Available: <https://doi.org/10.1007/s11263-020-01303-4>
- [36] H. Lin, S. Li, J. Xing, T. He, J. Yang, and Q. Wang, "High resolution aerosol optical depth retrieval over urban areas from Landsat-8 OLI images," *Atmospheric Environ.*, vol. 261, 2021, Art. no. 118591.
- [37] P. Su et al., "New particle formation event detection with mask R-CNN," *Atmospheric Chem. Phys.*, vol. 22, no. 2, pp. 1293–1309, 2022.
- [38] P. Su et al., "Retrieval of multiple atmospheric environmental parameters from images with deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 1005005.
- [39] R. W. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Eng.*, vol. 235, 2021, Art. no. 109435.
- [40] J.-J. Zhang, D.-X. Zhang, J.-N. Chen, L.-G. Pang, and D. Meng, "On the uncertainty principle of neural networks," 2022. [Online]. Available: <https://arxiv.org/abs/2205.01493>
- [41] Z. Zhang et al., "Multireceptive field: An adaptive path aggregation graph neural framework for hyperspectral image classification," *Expert Syst. Appl.*, vol. 217, 2023, Art. no. 119508.
- [42] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.
- [43] N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention Unet for building segmentation in remote sensing images," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–12, 2020.
- [44] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. 5th Int. Conf. Learn. Representations*, Toulon, France, 2017. [Online]. Available: <https://openreview.net/forum?id=Skq89Scxx>
- [45] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
- [46] D. Mutiibwa, S. Strachan, and T. Albright, "Land surface temperature and surface air temperature in complex terrain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4762–4774, Oct. 2015.
- [47] M. Schaap, A. Apituley, R. Timmermans, R. Koelemeijer, and G. De Leeuw, "Exploring the relation between aerosol optical depth and PM2.5 at Cabauw, the Netherlands," *Atmospheric Chem. Phys.*, vol. 9, no. 3, pp. 909–925, 2009.
- [48] Q. Pu and E.-H. Yoo, "Ground PM2.5 prediction using imputed MAIAC AOD with uncertainty quantification," *Environ. Pollut.*, vol. 274, 2021, Art. no. 116574.
- [49] J. Handschuh, F. Baier, T. Erbertseder, and M. Schaap, "Deriving ground-level PM2.5 concentrations over Germany from satellite column AOD for implementation in a regional air quality model," *Remote Sens. Clouds Atmos. XXV*, vol. 11531, pp. 5–16, 2020.
- [50] H. Hersbach et al., "The ERA5 global reanalysis," *Quart. J. Roy. Meteorological Soc.*, vol. 146, no. 730, pp. 1999–2049, 2020.



Peifeng Su received the Ph.D. degree in doctoral programme in atmospheric sciences from the University of Helsinki, Helsinki, Finland, in 2022.

His main research interests include deep learning, remote sensing, and computer vision.



Temesgen Abera receives the Ph.D. degree doctoral programme in geosciences from the University of Helsinki, Helsinki, Finland, in 2020.

Currently, he is a Postdoctoral Researcher with the Department of Geosciences and Geography, ECHOLAB, University of Helsinki. He is interested in using airborne and spaceborne observations, land surface flux modeling, and machine learning. His research interests include understanding and quantifying local and regional impacts of anthropogenic land surface changes on climate and ecosystem service in the tropics, particularly in East Africa.



Petri Pellikka obtains his Ph.D. in Physical geography from University of Oulu, Finland in 1998. He has been a Professor of Geoinformatics with the Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland, since 2002. He leads Earth Change Observation Laboratory and has a second affiliation with the Institute for Atmosphere and Earth System Research (INAR), University of Helsinki. He is interested in the causes and consequences of land change on climate, water resources, biodiversity, and livelihoods. Being the Director of

the Taita Research Station, University of Helsinki in Kenya, he focuses on environmental studies, especially in Kenya. His research interests include the application of remote sensing data and other geospatial data for land cover change detection with a focus on Africa.



Yanlong Guan received the Ph.D. degree renewable and clean energy from North China Electric Power University, Beijing, China, in 2021.

His main research interests include global environmental change and meteorological drought.