Intracity Temperature Estimation by Physics Informed Neural Network Using Modeled Forcing Meteorology and Multi-Spectral Satellite Imagery

Donghang Wu, Weiquan Liu, *Member, IEEE*, Bowen Fang, Linwei Chen, Yu Zang, Lei Zhao, Shenlong Wang, Cheng Wang, *Senior Member, IEEE*, José Marcato Junior, *Member, IEEE*, and Jonathan Li, *Senior Member, IEEE*

***Abstract*—Estimating urban surface temperature at high resolution is crucial for effective urban planning for climate-driven risks. This high-resolution surface temperature over broader scales can usually be obtained via satellite remote sensing for historical period. However, it can be hard for future predictions. This paper presents a Physics Informed Hierarchical Perception (PIHP) network, a novel approach for accurate, high-resolution and generalizable urban surface temperature estimation. The key to our approach is leveraging the implied temperature-related physics information of the land surface structure from high-resolution multi-spectral satellite images, thus achieving precise estimation or prediction for high spatial resolution urban surface temperature. Specifically, a semantic category histogram is first designed to describe the land surface structures. Based on this, a hierarchical urban surface perception network is proposed to capture the complex relationship between the underlying land surface features, upper atmosphere conditions and the intracity temperature. The proposed PIHP-Net makes it possible to generate models that can generalize across different cities, thus to estimating or predicting high-resolution urban surface temperature when the satellite land surface temperature (LST) observation is not available. Experiments over various cities in different climate regions in China show, for the first time, errors less than 2 Kelvin (for most of the cases) at the high resolution (60-by-60 meters grids), thus making it possible to predict future *intracity temperature* from forcing meteorology and multi-spectral satellite imagery.**

***Index Terms*—Land surface temperature, downscaling, multi-spectral satellite imagery, deep neural network.**

I. Introduction

Cities, as the hotspots of concentrated population and infrastructure, are where major climate-driven impacts occur [1], [2]. Effective urban planning and infrastructure-based growth strategies rely on high-resolution urban climate predictions [3], [4]. High-resolution and high-precision Land Surface Temperature (LST) prediction has been extremely challenging, especially over urban surfaces, because of their large heterogeneity [5], or the complicated natural and human behaviours, such as seasonal change, diurnal temperature differences, urbanization, population density, energy structure, etc [6], [7], [8], [9]. We acknowledge that ”high resolution” has different definitions for different fields. For example, sub-hundred meters resolution is typical in satellite remote sensing retrieval studies, and might not be considered as high resolution. However, in climate modeling studies, predicting intracity LST for future time horizon could be very challenging. Seasonal forecasts or long-term projections from climate models are either at coarse resolutions (>1 km) [3], [4] or completely missing urban landscapes [10]. The objective of this study is to establish an machine learning-based framework to predict intracity LST that can be used for urban climate applications using remote sensing data and deep neural networks, rather than LST retrieval. Therefore, the definition of ”high resolution” in this study means sub-hundred meters, aligned with urban climate studies.

Recent works [10], [11] employed the physics informed machine learning (PIML) paradigm to build an urban climate emulator to predict the citywide average temperatures on the global scale, demonstrating the large potential of incorporating physical understanding into machine learning for the urban temperature prediction. These PIML methods have also been demonstrated in modeling weather and climate processes[12] and other ecosystems[13]. Such a trend inspires us to deeply dig into the underlying physical reasons for the temperature prediction, not just considering it a sheer computer vision problem.

State-of-the-art methods to estimate LST can be mainly classified into two categories: physics-based methods and statistics-based methods. Physics-based models exploit the dynamic processes between the urban canopy and the atmospheric boundary layer to solve the temperature field [14], [15]. These methods are very computationally demanding and thus usually low-resolution (several kilometers and above). On

Manuscript received February 20, 2022. This work is supported in part by National Natural Science Foundation of China under Grant 61971363, and in part by China Postdoctoral Science Foundation under Grant 2021M690094, and in part by FuXiaQuan National Independent Innovation Demonstration Zone Collaborative Innovation Platform under Grant 3502ZCQXT2021003. *(D. Wu and W. Liu contribute equally to this paper.) (Corresponding author: Yu Zang, email: zangyu7@126.com)*

D. Wu, W. Liu, L. Chen, Y. Zang and C. Wang are with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen 361005, China (e-mail: donghangwu@stu.xmu.edu.cn; wqliu@xmu.edu.cn; willim@stu.xmu.edu.cn; zangyu7@126.com; cwang@xmu.edu.cn).

W. Bao and L. Zhao is with the Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA (e-mail: bowenf2@illinois.edu; leizhao.yale@gmail.com).

S. Wang is with the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA (e-mail: shenlong@illinois.edu).

J. Marcato Junior is with the Geomatics Laboratory, Faculty of Engineering, Architecture and Urbanism and Geography at the Federal University of Mato Grosso do Suln, Campo Grande, MS 79070-900, Brazil (e-mail: jose.marcato@ufms.br).

J. Li is with the GeoSTARS Laboratory, Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada (e-mail: junli@uwaterloo.ca).

the other hand, statistical methods seek to establish empirical relationships between the LST and other observational data, such as meteorological variables, land cover, geography, and vegetation indices [16], [17], [18], [19], [20], [21]. However, such methods are limited by the availability of the data and suffer from generalization ability and low accuracy [22].

Dynamic models leverage process-based equations that aim to resolve the physics within the urban canopy and atmospheric boundary layer to solve the temperature field. These models are highly computationally expensive and thus can hardly operate at a high spatial resolution, and have been limited by the physics represented in the model and the availability of urban surface characteristic datasets. For example, the dynamic downscaling methods using some widely-used models such as the Weather Research and Forecast (WRF) model [23] are usually conducted at 1-2 km resolution at the finest [4], [24], [25]; whereas the Computational Fluid Dynamics (CFD) based models are limited to very small scales such as a single urban block or street canyon [26], [27]. Their modeling accuracy is further subjected to the accuracy of the parameterization and representation of the physical processes in the models. These methods can hardly be applied for high-resolution urban temperature estimation, because of the infeasibility in resolving the very small physics scale.

Statistical models, on the other hand, seek to establish empirical relationships between LST and the auxiliary data, such as land cover, vegetation indices, and/or other observational data [28], [29], [30], [31]. Empirical at its core, traditional statistical downscaling methods are limited by: (i) the complexity of statistical methods used, (ii) the availability and reliability of the observed records, (iii) the relatively arbitrary choices of the features, and (iv) omission of the physics represented in the statistical models [32], [33], [34], [35]. These barriers significantly limit the traditional urban temperature estimation from generalization both spatially (i.e., upscaled to a larger region or applied to other study locations) and temporally (i.e., future forecast).

Recent efforts have started to explore the applications of both deep neural networks and physics-informed machine learning to tackle Earth and environmental science challenges. These applications pointed to some potentially promising avenues to address the aforementioned critical yet unresolved research gaps. Specifically, [36] modeled the lake temperatures across the depth and over time by combining physics-based models and deep learning methods. [37], [38] leveraged the advantages of convolutional neural network (CNN) in the processing of multi-channel images [39], [2], [40], applying CNN to sea surface temperatures maps and the oceanic heat content maps. These methods mainly targeted the LST prediction of homogeneous surfaces such as woodlands and waters, but nevertheless demonstrate a promising potential of the physics-informed neural networks paradigm. However, it is unknown whether this paradigm can successfully predict the intracity LST over urban surfaces which are largely heterogenous and with complex 3D structures.

In this study, we propose a Physics Informed Hierarchical Perception Network (PIHP-Net) to predict high-resolution urban LST directly from climate modeled forcing meteorology at a higher atmospheric level and land surface satellite imagery. Guided by the process-based physics understanding, such a network leverages the high-resolution multispectral satellite images, which are low-cost to acquire, to achieve accurate LST prediction at an high spatial resolution. Specifically, the semantic category histogram is first designed to describe the urban surface structure; then, the LST is decomposed by the bidimensional empirical mode decomposition to capture features at various scales. Based on this, a hierarchical urban surface perception scheme is proposed via a multi-branch network structure. This scheme captures the complex relationship between the land surface structures, upper atmosphere conditions and the intracity temperatures, and thus capable of generating accurate urban LST estimation results at a high resolution, when in the situation that LST cannot be generated in the future period or due to cloud cover.

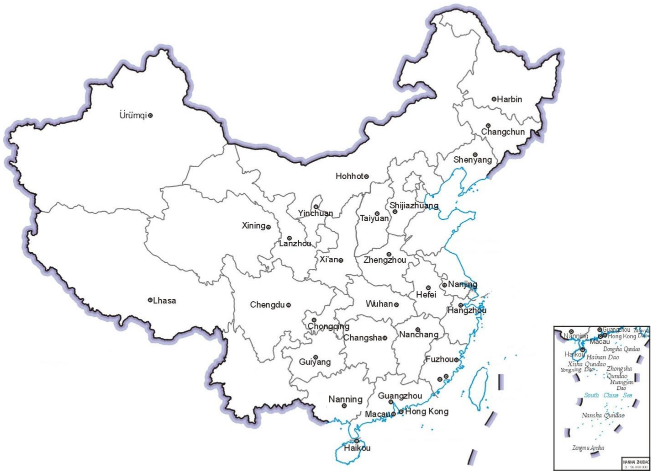


Fig. 1. Distribution of the selected urban in China.

II. Study area and data

In this work, 31 major or provincial capital cities in China are selected as the study area, see Fig. 1. The selected cities cover seven administrative regions of China and contain diverse types of terrain including metropolis, mountain, water, etc. We collected a suite of datasets from various sources in the study region to implement our PIHP-Net, described below. All of them are highly urbanized and populated. Research on these cities has great significance for assisting urban planning, carbon neutrality, and climate prediction. Specifically, four datasets are collected and processed in this paper, as described below:

The first dataset is the Landsat-based Ready-to-use (RTU) land surface temperature product from the CASEarth DataBank system[1], which is based on the USGS's Landsat8 OLI/TIRS sensor with a resolution of 30 meter. A single channel algorithm was used to retrieve LST in this product [41]. In this study, we collected the data from 2014 to 2019. All of the selected regions cover the urban area. Since we aim to estimate the block level urban temperature from the satellited image,

[1] http://databank.casearth.cn/

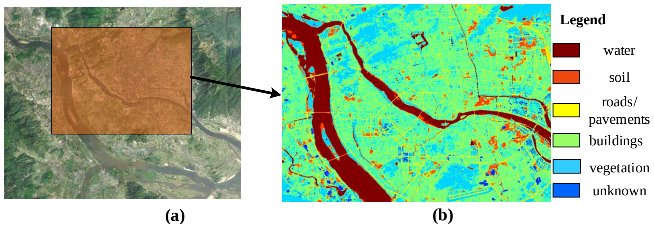


Fig. 2. Illustration of the generation process of GSCFM. (a) Satellite remote sensing images of Fuzhou, China. (b) Visualized map of the labeled. Inferred semantic categories (water, buildings, vegetation, soil, roads/pavements, and unknown) with text colored by the label color.

TABLE I

ATMOSPHERIC AND LOCATION DATA. THE AIR TEMPERATURE IN THIS STUDY MEANS THE ATMOSPHERIC TEMPERATURE AT THE REFERENCE HEIGHT (60M ABOVE THE SURFACE CANOPY TOP) IN REANALYSIS DATA OR CLIMATE MODELS.

Type	Name
Land Surface Forcings	Surface absorbed longwave radiation
	Surface income shortwave flux
Land Surface Diagnostics	Total precipitation land
Single-Level Diagnostics	Atmospheric temperature max
	Atmospheric temperature mean
	Atmospheric temperature min
Analyzed Meteorological Fields	Surface pressure
	Atmospheric temperature at the reference height
	Eastward wind
	Northward wind
	Specific humidity

thus the 60m-by-60m of the urban grid has been divided in our experiments. Thus, the corresponding LST data has also been aggregated to 60m resolution. An example visualized map is shown in Fig. 5 (a).

The second dataset is the Nationwide remote sensing images of the major and provincial capital cities in China, along with the semantic label. Fig. 2 (a) and (b) show the examples of second dataset of Fuzhou, China, in which (b) is the visualized result of the semantic label, with the different urban surface properties denoting as various colors. These multi-spectral remote sensing images have a resolution of 1 meter from the Google Earth website [2] for all major provinces and key cities in China. Besides the spectral information, we find that the LST is also highly correlated with the properties of the underlying surface. So a pixelwised labeling of the remote sensing image is also involved in this dataset. Specifically, each pixel is automatically labeled with predefined categories (including water, buildings, vegetation, soil, and roads/pavements, etc.) by the state-of-art classification technique [42] and the accuracy of the classification is 96.5%.

The third dataset is the temperature measured a by the weather station from the website of China's National Greenhouse Data System[3]. The dataset contains temperatures measured at weather stations in urban centers in 30 cities for each day from 2014-2019. And the kind of temperatures used is daily average air temperature at 2m.

The fourth dataset is the Normalized Difference Vegetation Index(NDVI) of the Landsat-based Ready To Use (RTU) products from the CASEarth DataBank system[4], which is based on the USGS's Landsat8 OLI/TIRS sensor with a resolution of 30 meter. And Normalized Difference Built-up Index (NDBI), for Landsat 8 data from USGS Global Visualization Viewer website[5], NDBI = (Band 6 – Band 5) / (Band 6 + Band5)[43]. Meanwhile, all of the selected regions are the same to first dataset.

The last dataset is the atmospheric forcing data, provided by the NASA MERRA-2 reanalysis data system [44]. These data are publicly accessible from the NASA MERRA-2 website[6]. The resolution of such data is 0.5^o latitude $\times$ 0.625^o longitude. The atmospheric forcing data have a temporal resolution of days and reflect the overall weather conditions of the city on that day. A list of these variables is shown in Table I.

Due to the limitation of national policy, the raw high-resolution multispectral satellite remote sensing images are not open source. As an alternative, we open-source the calculated Geographical Semantic Category Fraction Matrix (GSCFM, details can be viewed in the Methods section) of all cities. The complete dataset will be released via the FTP server later.

III. METHODS

Considering the temperature prediction as a simple image-based regression problem, the network may be confused by the complex image pattern. Thus, the relationship between the underlying surface, upper air condition, and the LST may fail to be captured by the network.

We propose a physics-informed hierarchical urban surface temperature perception scheme to guide the overall estimation process by related physical factors. First, by employing previous image ground targets classification network [42], the satellite imagery is first assigned a pixel-wise semantic label according to the geophysical categories. In this work, 11 urban surface categories (cultivated land, garden land, forest land, grassland, buildings, roads, structures, excavated land, bare land, water, unknown or covered with snow) are applied. Finally, the 11 categories are divided into 6 major categories. Second, we design a grid-wised (60-by-60 meters) semantic category histogram to further aggregate the urban surface features. Such a dense urban surface descriptor, and the atmospheric forcing and weather station observation data are fed into the network. Then, by passing a proposed bidimensional empirical mode decomposition-based hierarchical network, the mapping relationship between the urban surface temperature and the input data can be captured at different scales, producing the estimated temperature for each grid of the testing city.

A flowchart of predicted LST is shown in Fig. 3. Implementation of the whole is performed by three stages: 1) data spatial matching and preprocessing; 2) prediction of LST; and 3) correction and generation of LST. In the stage of data spatial matching and preprocessing, labeled remote sensing image, NDVI, NDBI, and atmospheric forcing are matched under the same space. For NDVI, NDBI, and atmospheric forcing are

[2] https://earth.google.com/
[3] http://data.sheshiyuanyi.com/WeatherData/
[4] http://databank.casearth.cn/
[5] https://glovis.usgs.gov/
[6] https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/

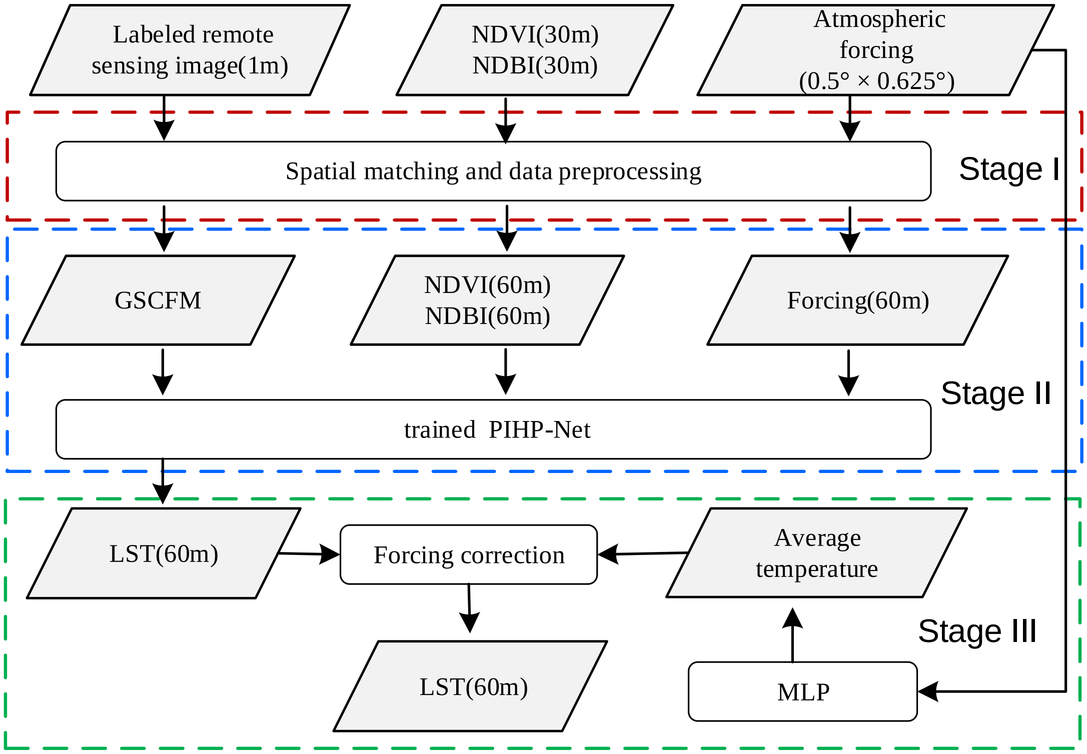


Fig. 3. Flowchart of the proposed method for predicting the 60-m LST from labeled remote sensing image, NDVI, NDBI, and atmospheric forcing data.

mapped onto each 60m grid, different scales of GSCFM are constructed as input to PIHP-Net. In the second stage, LST is predicted through trained PIHP-Net. Finally, in the stage of correction and generation of LST, the LST from stage II goes through forcing correction to obtain a 60m precision LST. Details of forcing correction are given in Section III-B3.

A. Physics informed hierarchical perception scheme

1) Geographical semantic category fraction matrix: To introduce the structural distribution information of local city embedded in remotely sensed image into neural network, a straightforward approach is to import the entire image directly into the network. However, we found that this approach has difficulty capturing the general relationship between temperature and land surface information, because the high complexity of the urban surface features leads to poor generalization of the whole system. So we designed a semantic category histogram to summarize the potential factors that influence the local-scale urban surface temperature. This descriptor is subsequently denoted as the Geographical Semantic Category Fraction Matrix (GSCFM).

The GSCFM contains the ground structure information from labeled remote sensing images with a spatial resolution of 1-by-1 m. The ground structure information is set to 5-by-5 m as one cell and calculated from the labeled images within each cell, as described below. Each cell is described by a high-dimensional feature, which implies the underlying surface characteristics that affect the local temperature. In our approach, the five main temperature-related structure categories (including water, buildings, vegetation, soil, and roads/pavements) are summarized by previous semantic labeling works, with various colors in the dots indicating different urban surface properties. The urban LST data were obtained from NASA's Landsat satellite measurements with a spatial resolution of 60-by-60 m. For each 60-by-60 m grid, we build a larger matrix by a batch of 5-by-5 meter cells. The matrix size is $M \times M \times d$, where M denotes the number of cells and d denotes the dimension of the feature vector of one cell. Each category of GSCFM in one cell is defined by the following equation:

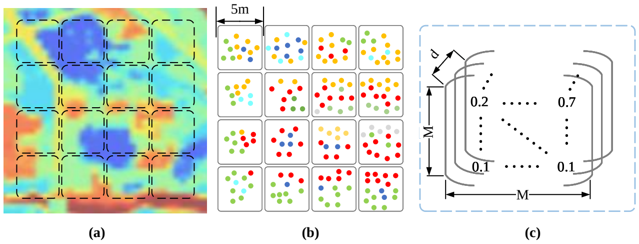


Fig. 4. The illustration of the generation process of GSCFM. (a) The background is part of the land surface semantic label image, and a dummy black box represents the 5-by-5 meter cells. (b) Visualized result of different categories in every cell, different colors represent different categories. (c) For each cell, we can calculate the percentage of each category in one cell. There are d categories in total, so each cell corresponds to a vector of $1 \times d$. And for a 60-by-60 meter grid, we can select $M \times M$ cells to describe the surface structure and domain information of this gird, and this is one example of the GSCFM with size $M \times M \times d$.

$$d(l)=\frac{C\left(S_{l}\right)}{C(S)} \tag{1}$$

where S denotes the set of pixels on a particular cell, $C(\cdot)$ denotes the number of points in S, S_l denotes the number of pixels with category l in S. Fig. 4 shows examples of the constructed GSCFM.

2) Multi-scale hierarchical urban surface perception: Due to the complexity of the surface structure and the fact that the temperature of a single 60-by-60m grid may be influenced by the surrounding surface structure, the network is hard to capture the relationship between the GSCFM and LST directly. Inspired by previous signal processing works [45], [46], [47] and to reduce the complexity of the Non-linear transformation in the neural network, we propose a bidimensional empirical mode decomposition (BEMD) to decompose raw data into multi-scales. Empirical mode decomposition (EMD) is a classical signal analysis tool [48] that decomposes data into a series of low frequency bases, which are called 'intrinsic mode functions (IMF)'. The BEMD algorithm treats the image as a signal and decomposes it to obtain different IMFs, and each IMF contains the information of different scales. Thus, our idea is to build the connections between the IMFs of the LST image and the GSCFM at different scales. Such a multi-scale perception scheme allows the network to capture the relationship between the urban surface and the temperature in various scales, leading to better estimation results.

Consider the LST data as a one-channel image, denoted as $I(\mathbf{p})$. Then, the decomposition process can be formally written as:

$$I(\mathbf{p})=h(\mathbf{p})+r(\mathbf{p}), \tag{2}$$

Here, $h(\cdot)$ is the intrinsic mode representing fine-scale distribution information. $r(\cdot)$ is the residue layer obtained by averaging the envelopes of local maxima $E(\cdot)$ and minima $e(\cdot)$, i.e. $r(\mathbf{p}) = (E(\mathbf{p}) + e(\mathbf{p}))/2$. For the computation and localization of local extrema, we use directly finding the extrema from each sub-square neighborhood. In addition, we simply use uniform cubic spline interpolation to compute the envelopes $E(\cdot)$ of maxima and $e(\cdot)$ of minima from the local extrema. Such a decomposition can be recursively applied to obtain multi-scale intrinsic mode functions, as written by:

$$I(\mathbf{p})=\sum_{i=1}^{n} h_{i}(\mathbf{p})+r_{n}(\mathbf{p}). \tag{3}$$

Hear, n denotes the number of recursions. As shown in Fig. 5, three decomposition operations are applied, where Fig. 5(a) is the original LST image. Fig. 5(b)-(d) are the acquired intrinsic mode functions that reflect the temperature signals in different scales. Fig. 5(e) is the residual image after 3 iterations. These modes along with the residual are connected with different branches of the network, to guide the GSCFM in different spatial scales.

B. Network Architecture

The proposed PIHP-Net, as shown in Fig. 6, consists of two parts, multi-scales encoder and information parallel decoder.

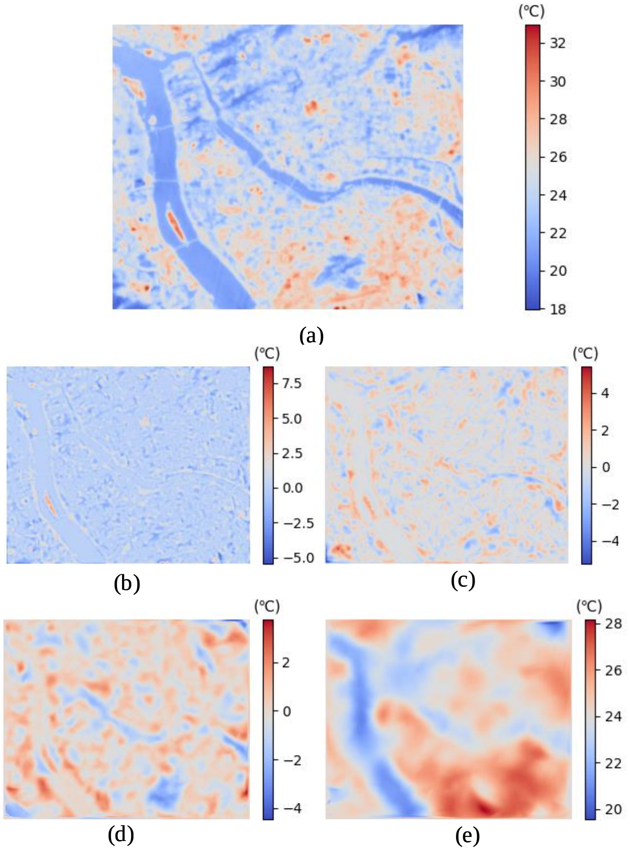


Fig. 5. Illustration of the result of BEMD, each colorbar's number represents the temperature in Celsius (°C). (a) LST of Fuzhou, China. (b)-(d) are intrinsic mode functions that reflecting the temperature signals in different level of detail information. (e) is the residual term from BEMD.

1) Multi-scales encoder: For the sake of multi-scale GSCFM data, accordingly we apply the multi-scale encoder to them. The input of our multi-scales encoder is varied spatial scale GSCFM (control by different M), NDVI, NDBI, (match to the GSCFM) and forcing. Here we set $M = 35 + 10 * i, i = 1, ..., n$ to obtain different scales GSCFM, then leveraging Local-surface feature extractor to encode the feature from them, respectively.

After the encoding operation, the local surface feature information into latent vector F_i, for $i = 1, ..., n$. All F_i are then concatenated, forming a latent feature map P, with size of $1024 \times n$. Next, the multilayer perceptron (MLP) is applied to integrate the latent vector P into the final latent vector F_0, size of $F_0 = 1024$, which contains different scales of GSCFM information. A batch of independent Forcing feature extractor will be used to extract the large-scale atmospheric feature G_i, for $i = 0, ..., n$. For all independent channel, we concatenate F_i and G_i to generate H_i, which contains the information of broader-scale atmospheric forcing factors and high-resolution local urban surface feature information.

The details of Local-surface feature extractor and Forcing feature extractor are as follows:

(i) Local-surface feature extractor

In order to dig the structural information of the urban surface, we first introduce ResNet [39], which aims to extract

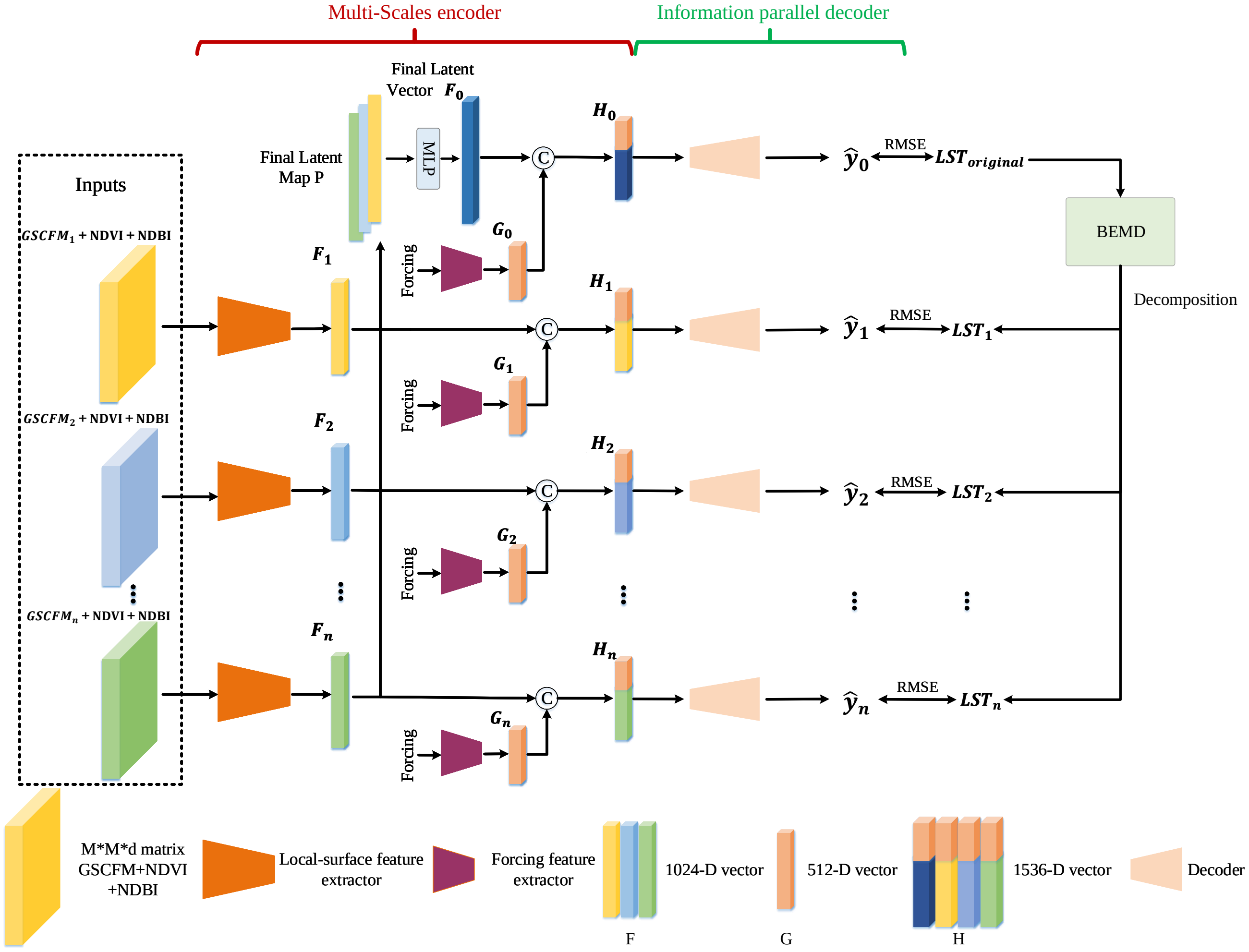


Fig. 6. An overview of the proposed physics-informed hierarchical perception network (PIHP-Net).

local surface features and spatial structure information from the proposed GSCFM. This branch enables the neural network to capture high-resolution variations in the temperature of the urban surface. This network consists of a 5-stage deep residual network, each stage containing two residual blocks. The first block is composed by two 3×3 convolutional layers sequentially. The stride is set as 2 and 1, respectively, and a skip connection is used to align the output shapes. The second block consists of two convolutional layers, where the stride is set as 1. The last stage consists of a 3×3 convolution operator, which adjusts the output size instead of a pooling operation. Batch normalization and ReLU layers are applied after each convolutional layer.

(ii) Forcing feature extractor

Considering the impacts of atmospheric states on the LST, we employ a MLP to encode the primary atmospheric forcing variables from the physics-based climate model. This procedure can be considered as a deep learning 'solver' of the physical equations in those process-based models, in a way to mimic the dynamic simulations [10]. For the MLP, we introduce five layers to encode the feature vector into a vector G, with size 256. SeLU [49] is used to avoid gradient explosion and vanishing.

2) Information parallel decoder: In order to decode the aggregated feature H_i to predict LST, we fed them into regression branch, respectively. Each branch is comprised of three fully-connected layers. Next, each H_i is mapped to temperature $\hat{Y}_i$ by the corresponding regression branch. Following the previous work [36], we employ the root mean squared error (RMSE) and the L_2 normalization of the network weights to measure the loss.

3) Forcing correction: In the whole framework, the upper-level atmosphere condition is related to the forcing data, in which humidity, radiation, precipitation and the temperature of the upper atmosphere are involved. However, due to the fact that the LST is actually related to the near-surface temperature, there is an intrinsic deviation between the mean value of LST and the forcing data. To demonstrate this, we collect about 40 pieces of data over different seasons for 30 major cities under different climate regions in China. Then the mean value of the underlying surface in the LST data, along with mean value of upper air in the corresponding forcing data are applied

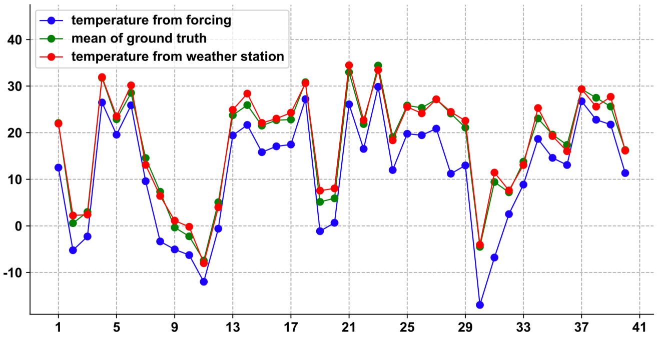


Fig. 7. Comparing the temperature from weather stations, forcing and average LST. The ordinate represents the temperature in Celsius (°C).. The abscissa represents the city number.

to generate a scatter plot, as shown in Fig. 7.

In Fig. 7, the average 6.38K distance between the LST and forcing can be observed, and such bias may largely affect the prediction result. To address this issue, we initially planed to find the mapping relationship between the mean value of LST and forcing by a simple branch in the network. However, limited by the period of satellite, the LST data are sparse and thus difficult to train the branch. Therefore, in our approach, the data from the weather station are included as the intermediate. The weather station data is dense and easy to acquire for cities, and most importantly, it matches really well with the mean value of LST. As shown in the red line in Fig. 7, the average bias of 0.42K to the LST can be observed. Such an observation inspires us to fit the mean value of weather station by the input forcing, thus to correct the upper air forcing closer to the near-surface cases.

Specifically, the branch is simple: the mean value of LST is predicted by leveraging the MLP to fit the weather station data, and then the middle result is used as the input of the decoder network.

4) Loss function.: Following the previous work [36], we employ the root mean squared error (RMSE) and the L_2 normalization of the network weights to measure the loss. Each H_i, for $i = 0, ..., n$, has independent loss $Loss_i$, which is used to guide the training of the network, and the final loss L is the sum of them. The overall loss can be written as:

$$L = \sum_{i=0}^{n} Loss_i \tag{4}$$

$$Loss_i = \underset{\mathbf{W},b}{\operatorname{argmin}} L(LST_i, \hat{Y}_i) + \lambda R(\mathbf{W}) \tag{5}$$

$$L(Y, \hat{Y}) = \frac{1}{N} \sum_{j=1}^{N} (y_j - \hat{y}_j) \tag{6}$$

$$R(\mathbf{W}) = \|\mathbf{W}\|_2 \tag{7}$$

Here, Y_i, for $i = 1, ..., n-1$ is LST_i, which is generated by BEMD, and Y_0, Y_n represent the ground truth set. $\hat{Y}_i$, for $i = 0, ..., n$ represent the predicted results set from H_i. $\mathbf{W}$ and b are the combined coefficients of weights and bias terms, $y_j \in Y, \hat{y}_j \in \hat{Y}, j = 1, ..., N$, N is the data size of $Y, \hat{Y}$, and λ is the weight of the regularization term.

IV. Results and Discussion

A. Implementation details and baselines

1) Implementation details: The models are evaluated quantitatively on test sets based on the root mean square error (RMSE), defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / n} \tag{8}$$

We implement our PIHP-Net on PyTorch [50]. The Network uses end-to-end training mode. In the training phase, we adopt Adam solver [51] with an initial learning rate 0.001, which decayed by $0.5^{1/500}$for each epoch.

2) Baselines: To our knowledge, the proposed PIHP-Net is the first physics informed deep neural network, which attempt to estimate the urban temperature at an high resolution (60-by-60 meters). In consideration the huge calculated amount of the dynamic equation based methods at such resolution, our comparisons are mainly focused on previous statistic model based methods, such as linear regression [52], K-Nearest Neighbors (KNN) regression [53], random forest regression [54], and deep neural network-based method, such as ResNet [39]. All the statistic model-based methods are implemented based on Scikit-learn [55], specifically, the parameter settings of these methods are as follows:

- **Linear Reg** uses linear regression for simulation. We concatenate GSCFM and forcing as input. Specifically, each dimension of the $M \times M \times d$ matrices is averaged to a specific value, and the original matrix is reshaped to a $1 \times d$ vector. This vector is then imported into various regression methods. That is, linear models take the form:

$$y = \sum_i \beta_i x_i + \varepsilon \tag{9}$$

  where y is the response (LST), x_i is GSCFM and forcing reshaped to a $1 \times d$ vector, β_i is how the LST changes linearly with each x_i, and ε is the normally distributed error.
- **KNN Reg** is KNN regression. Specifically, the number of neighbours in the KNN regression is set as 4. The maximum depth of the tree is set as 30. The input feature is the same as Linear Reg.
- **RF Reg** is random forest regression. The number of trees in the random forests regression is set as 150.
- **ResNet** is based on ResNet50, and the corresponding parameters are the same as the ResNet in our network.

B. Ablation study and sensitive analysis

In this section, we design a series of experiments to evaluate the effects of the various components in the proposed PIHP-Net.

1) Ablation of hierarchical urban surface perception scheme: The motivation of the proposed BEMD based hierarchical perception scheme is that PIHP-Net senses the surface structure information at different scales. The hierarchical perception scheme helps network more accurately simulate the surface temperature of the city. A group of experiments is

TABLE II
THE AVERAGE TEMPERATURE RMSE OF ABLATION OF HIERARCHICAL URBAN SURFACE PERCEPTION SCHEME

	Guangzhou	Zhangzhou	Shenyang	Avg. Error (Kelvin)
PIHP-Net^0	0.62	0.64	0.66	0.64
PIHP-Net^1	0.58	0.58	0.62	0.59
PIHP-Net^2	0.56	0.56	0.59	0.57
PIHP-Net^3	**0.52**	**0.53**	**0.55**	**0.53**
PIHP-Net^4	0.56	0.59	0.63	0.59

TABLE III
THE RMSE FOR DIFFERENT LAND COVERS OF PIHP-NET^3 IN THE ABLATION OF HIERARCHICAL URBAN SURFACE PERCEPTION SCHEME.

	Guangzhou	Zhangzhou	Shenyang	Avg. Error (Kelvin)
water	0.32	0.36	0.39	0.36
soil	0.41	0.45	0.49	0.45
roads/pavements	0.58	0.61	0.54	0.58
buildings	0.62	0.63	0.60	0.62
vegetation	0.45	0.49	0.51	0.54
unknown	0.71	0.66	0.75	0.71
total	0.52	0.53	0.55	0.53

TABLE IV
THE ABLATION OF HOW THE AGGREGATE BRANCH AFFECT THE RESULTS

	Guangzhou	Zhangzhou	Shenyang	Avg. Error(Kelvin)
PIHP^3 without aggregate branch	0.62	0.65	0.70	0.66
PIHP^3 with aggregate branch	**0.52**	**0.53**	**0.55**	**0.53**

TABLE V
THE AVERAGE TEMPERATURE RMSE OF SENSITIVE ANALYSIS OF PIHP NETWORK FOR INPUT DATA ERROR

	Guangzhou	Zhangzhou	Shenyang	Avg. Error (Kelvin)
without error	0.52	0.53	0.55	0.53
5% error of labeled remote sensing image	0.56	0.58	0.61	0.58
10% error of labeled remote sensing image	0.59	0.61	0.64	0.61
15% error of labeled remote sensing image	0.62	0.65	0.70	0.66
10% error of LST	0.59	0.61	0.61	0.60
20% error of LST	0.65	0.63	0.71	0.66
30% error of LST	0.71	0.78	0.82	0.77

designed to demonstrate the effectiveness of such a scheme, and confirm the appropriate layer number.

We selected 12 pieces of data, for the cities Guangzhou, Zhangzhou, and Shenyang in China, between January 1, 2015 and December 31, 2019. These data are selected to cover different seasons of the year, and to ensure the reliability of the evaluation results. All the data are handled as cloud-free to avoid aberrant temperature sampling by satellites. For each piece of data, the 70% area is chosen as the training data, and the rest 30% is applied for testing.

In the experiment, the average RMSE of the data for the PIHP-Net with different numbers of the layer are collected, and the corresponding results are shown in Table II. Where PIHP-Net^0 denotes the network without hierarchy, i.e. with number of branch 1. PIHP-Net^1, PIHP-Net^2, PIHP-Net^3 and PIHP-Net^4 correspond to the number of the perception branch $2-5$. The results demonstrate that the hierarchical perception scheme brings about 6% performance increasing (the average temperature RMSE decreased from 0.64 to 0.59). With the increasing number of perception branch, the best performance appears at PIHP-Net^3, corresponding to 4 branches, and the RMSE increased 17% in total. Such results demonstrate that, the proposed hierarchical perception scheme makes the PIHP-Net better capture the nonlinear relationship between the LST and the urban surface structures. In addition, the average RMSE of the data for the PIHP-Net^3 with different land covers are collected, and the corresponding results are shown in Table III. Except for unknown, the average RMSE of roads/pavements and buildings are higher at 0.62 K and 0.58 K, respectively. The average RMSE of water is the lowest, is 0.36 K. This difference may be due to the fact that the temperature of water varies less, while the temperature of roads/pavements and buildings are more sensitive to environmental as well as human factors.

2) Ablation of the aggregate layer: As shown in Fig. 6, in our proposed PIHP-Net, an aggregate branch of the original LST data is also involved. In the experiments, we observe that the original LST image contributes significantly to the prediction results. In order to verify this, an ablation study is designed. Whereas the experiment data and process are the same as the Section IV-B1, the number of hierarchical layers is set as 4, and the results with and without the aggregate branch are collected, as listed in Table IV. The average RMSE of various cities is shown separately; the overall performance gains approximately 19% increasing for different cities. The underlying reason of this improvement is likely that the aggregate branch can correct the error in the LST edge part of the decomposition by BEMD.

3) Sensitive analysis of PIHP network: In our proposed PIHP-Net, the input training data mainly consists of GSCFM constructed from the labeled remote sensing images as well as LST ground truth. To analyze the sensitivity of the network to input data errors, the sensitive analysis to them is conducted as follows:

The sensitive analysis experiments are based on the ablation studies in Table II(where the same data and cities are selected). We have added the noise to 5%, 10%, and 15% of the real data by random sampling. Where the noise has been added to two sources of data (labele of remote sensing images and the LST ground truth) separately. For the LST data, where the noise of ±2K has been added to the clean data. The results of different percentage errors are collected, as listed in Table V.

For the labeled remote sensing images, the RMSEs under 5%, 10%, and 15% percent of noise, increase 9%, 15%, 23% respectively. For the LST data, the RMSEs are 13%, 24%, 44% higher, respectively.

The supplemented experiments show that the error of the input training data has a significant impact on the experimental results, and the results are more sensitive to the errors of the LST than the label of remote sensing images.

C. Comparisons

To comprehensively compare the proposed approach and previous statistical models and deep neural network-based methods, two additional groups of experiments are designed. Specifically, the first group of experiments focuses on the single city evaluation to verify the performance for estimating

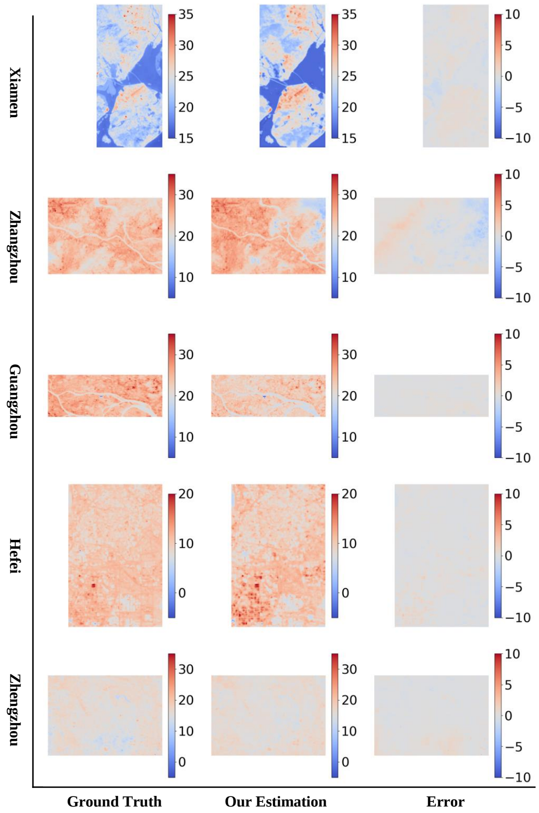


Fig. 8. The experiment of Single city temperature prediction visualization plots on Xiamen, Zhangzhou, Guangzhou, Hefei and Zhengzhou, China. Each colorbar's number represents the temperature in Celsius (°C). The vertical coordinates represent the city names and the horizontal coordinates represent the truth value, our simulation and the error, respectively. All visualization plots are averaged results from the corresponding multiple data.

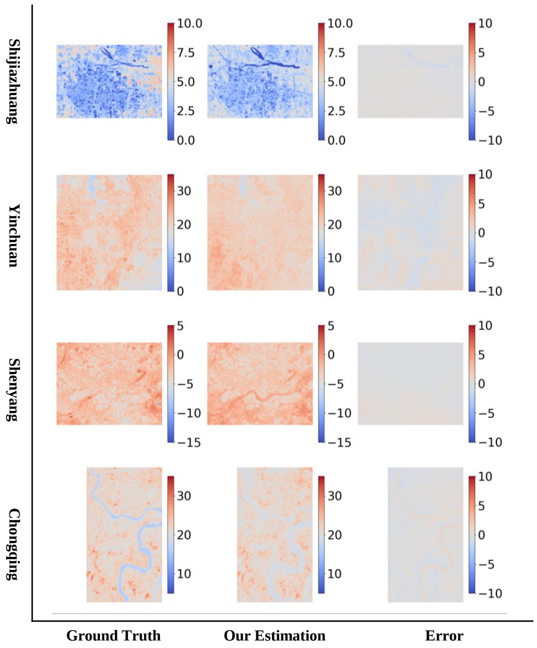


Fig. 9. Experimental single-city temperature prediction visualization plots on Shijiazhuang, Yinchuan, Shenyang and Chongqing.

a certain city at different times. The second group of experiments, on the other hand, pays more attention to the large scale evaluation, where most of the major cities in China are involved is showing the fidelity of the proposed PIHP-Net.

1) Single city temperature prediction: We conducted this experiment in 9 key cities distributed in major regions of China. Data from a single city at multiple time points are applied for training, and the temperatures at other time points are applied for testing. Because the surface structure of a city barely changes in $1-2$ years, while the atmospheric condition changes dramatically and is closely related to the seasons. Thus, for each city, $6-8$ pieces of data over four seasons are chosen for training, and $2-4$ pieces of data are applied for testing.

Methods of linear regression, KNN regression, and random forest regression (denoted as Linear Reg, KNN Reg and RF Reg separately) and the previous deep network ResNet are employed for comparison. The proposed GSCFM is not included for these baselines because we aim to figure out how the physics-informed GSCFM affects the results. The corresponding parameter settings are mentioned in Section IV-A2. The comparison without the hierarchical perception scheme is also involved, denoted as PIHP-Net^0.

The average RMSE for each city is collected, with corresponding results listed in Table VI and the average errors over all nine cities for each method are shown in the last column. It is viewed that, previous statistic model-based approaches get the average error more than 3.3K over all the cities. The linear model got the worst results with 4.69K, because the relationship between the LST, underlying surface, and upper atmosphere is complex and nonlinear. The results of KNN and random forest regression are similar, and appear unstable over different cities (with about 3.33K for the best case and 5.20K for the worst case). Thanks to the proposed GSCFM and the hierarchical perception scheme, our approach obtains an average error of 1.83K for all cases. It is worth noting that for some of the cities such as Zhangzhou, Guangzhou and Shijiazhuang, the prediction error is close to the typical observational error of the satellite (e.g., about 1K). The average error of ResNet (i.e. without the physics information GSCFM and hierarchical perception scheme), is 2.36K, i.e., 22% higher than our approach. The average error of PIHP-Net^0 is 2.24K, i.e., 18% higher. A visualization of the results for the nine cities is shown in Fig. 8 and Fig. 9.

2) "Unseen" city temperature prediction: In this part, we attempt to evaluate the ability of PIHP-Net to predict the temperature of an unseen city in the training set. Such an experiment includes almost all the major cities in China. Considering the large area of China, this group of experiments is developed following the official major regions division of China, which are South, Southwest, Northwest, Northeast, Central, North, and East.

For each region, $2-4$ cities, and 4 pieces of data for each city of over different seasons are chosen for training, as shown in Table VII. Then, the remaining $1-2$ cities with $3-6$ pieces of data are applied for testing. Similarly, methods of Linear Reg, KNN Reg, RF Reg, ResNet and PIHP-Net^0 are employed for comparison. The average errors over all testing

TABLE VI
OVERALL PERFORMANCE COMPARISON OF DIFFERENT APPROACHES ON THE EXPERIMENT OF SINGLE CITY FUTURE TEMPERATURE PREDICTION. A SMALLER VALUE INDICATES A BETTER PERFORMANCE. THE AVERAGE ERRORS OVER ALL THE NINE CITIES OF EACH METHOD ARE SHOWN IN THE LAST COLUMN

	RMSE (Kelvin)									
Model	Xiamen	Zhangzhou	Guangzhou	Hefei	Zhengzhou	Shijiazhuang	Yinchuan	Shenyang	Chongqing	Avg. Error
Linear Reg	4.61	4.26	5.05	4.50	5.63	4.44	6.23	3.37	4.13	4.69
KNN Reg	3.99	3.78	3.33	3.41	4.14	3.59	5.20	4.50	3.98	3.99
RF Reg	3.84	3.40	3.37	3.48	4.07	3.36	4.93	3.79	3.54	3.75
ResNet	2.72	2.18	1.75	1.81	2.63	2.32	3.25	2.10	2.47	2.36
PIHP-Net^0	2.60	2.03	1.65	1.79	2.57	2.20	3.06	1.96	2.34	2.24
PIHP-Net	**2.00**	**1.78**	**1.32**	**1.44**	**2.13**	**1.78**	**2.52**	**1.66**	**1.87**	**1.83**

TABLE VII
SPLIT OF TESTING AND TRAINING CITIES FOR EACH REGION

Region	City for training	City for testing
South China	Nanning,Guangzhou,Haikou,Hong Kong	Macao
Southwest China	Guiyang,Chongqing,Chengdu	Lhasa
Northwest China	Xining,Yinchuan,Urumqi	Xi'an,Lanzhou
Northeast China	Changchun,Harbin	Shenyang
Central China	Wuhan,Changsha	Zhengzhou
North China	Shijiazhuang,Hohhot	Taiyuan
East China	Nanchang,Nanjing,Jinan	Hangzhou,Fuzhou

TABLE VIII
COMPARISON OF RMSE ON SEVEN REGIONS FOR TEMPERATURE PREDICTION OF TESTING CITIES USING PIHP-NET AND OTHER BASELINE METHODS. THE AVERAGE ERRORS OVER ALL TESTING CITIES OF EACH METHOD ARE SHOWN IN THE LAST COLUMN

	RMSE (Kelvin)									
	South China	Southwest China	Northwest China		Northeast China	Central China	North China	East China		Avg. Error
Model	Macao	Lhasa	Xi'an	Lanzhou	Shenyang	Zhengzhou	Taiyuan	Hangzhou	Fuzhou	
Linear Reg	5.06	7.03	6.84	6.30	5.22	5.70	3.49	5.66	6.83	5.7
KNN Reg	4.00	5.93	5.55	4.90	3.72	4.54	2.82	4.64	5.46	4.62
RF Reg	4.13	5.88	5.28	5.19	3.63	4.26	3.33	4.33	5.24	4.58
ResNet	2.20	4.44	4.20	3.66	2.03	2.52	1.74	2.44	2.60	2.87
PIHP-Net^0	1.91	3.93	3.76	3.42	1.83	2.25	1.54	2.23	2.42	2.59
PIHP-Net	**1.80**	**3.77**	**3.46**	**3.17**	**1.72**	**2.10**	**1.43**	**2.09**	**2.26**	**2.42**

cities of each method are collected, as shown in Table VIII. The results show that, as the testing scale increases, the overall prediction error suffers, decreasing in different degrees for various methods. Specifically, results of previous statistical-based works remain unstable, maximum prediction error is more than 7.03K at the city of Lhasa, and the average error over all the test cities also exceeds 4K.

For the deep networks, average errors of 2.87K, 2.59K and 2.42K for ResNet, PIHP-Net^0 and our approach, respectively, are observed. As shown in Fig. 10 and Fig. 11, due to the more complicated underlying urban surface and the vagaries of climate, the advantage of the proposed PIHP is still demonstrated but not as substantial as compared to the single city prediction.

D. Discussion

In this section, we focus mainly on the error analysis of the experimental results, where some of the interesting factors besides the underlying surface and atmospheric forcing conditions are observed to react on the intracity LST. Thus, in this section, some of the instructive experimental results are presented, aiming to inspire more accurate intracity LST estimation systems in the future.

1) Errors from regional climate: As shown in Fig. 10 and Table VIII, it is worth to noting that, for two regions, Southwest and Northwest China, the prediction error for cities Lhasa, Xi'an, and Lanzhou significantly exceed the average. To find out the potential reasons, we first view the administrative division of China, as shown in Fig. 12. The Southwest and Northwest cover the largest area, about 50^o longitude and 39^o latitude, significantly larger than other regions. Such a large geographical span brings vagaries of climates for these two areas.

For the Southwest, the testing city Lhasa is located on the Tibet Plateau, with an average altitude of 3650m, while the training cities of this area Guiyang, Chengdu, and Chongqing, are all located at the Szechwan Basin, with an altitude of 250m. Such geographical conditions bring extremely differential climate features for the training and testing cities, thus leading to the poor estimation performance.

For the Northwest, a similar situation is observed. Where Urumqi locates near the Junggar Basin, with desert all around. Xining locates on the Qinghai plateau, with an average altitude of 2200m. Yichuan locates in the Helan mountain area. Lanzhou locates at Yellow River Valley basin, on the Loess Plateau. Xi'an locate at Kuan-chung Plain, with 8 river systems around. All the five major cities in Northwest China have their own unique geographical and climatic characteristics, leading

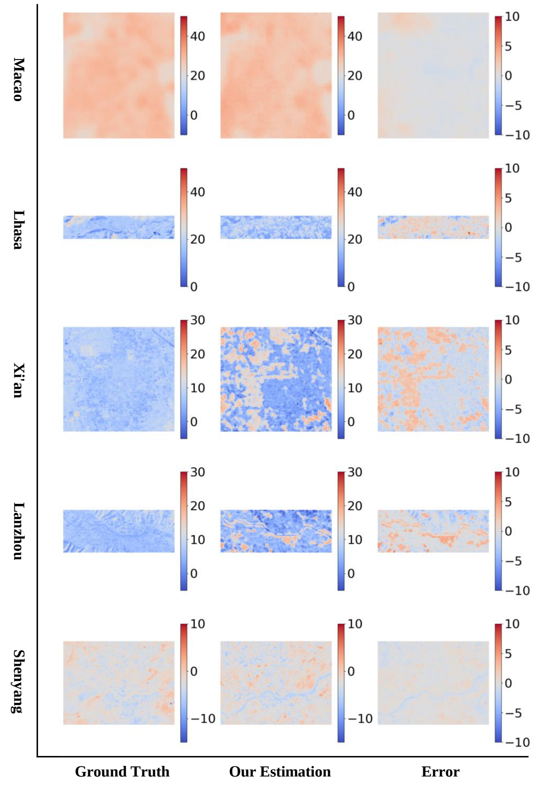


Fig. 10. Experimental "Unseen" city temperature prediction visualization plots on Macao, Lhasa, Xi'an, Lanzhou and Shenyang.

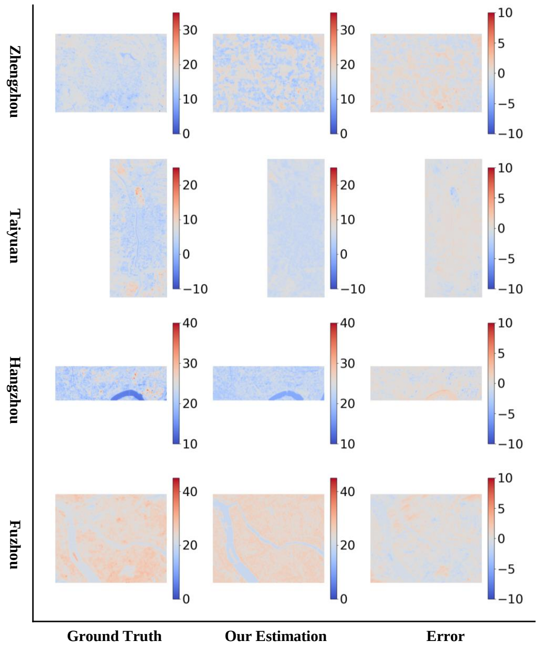


Fig. 11. The "Unseen" city temperature experiment prediction visualization plots on Zhengzhou, Taiyuan, Hangzhou and Fuzhou.

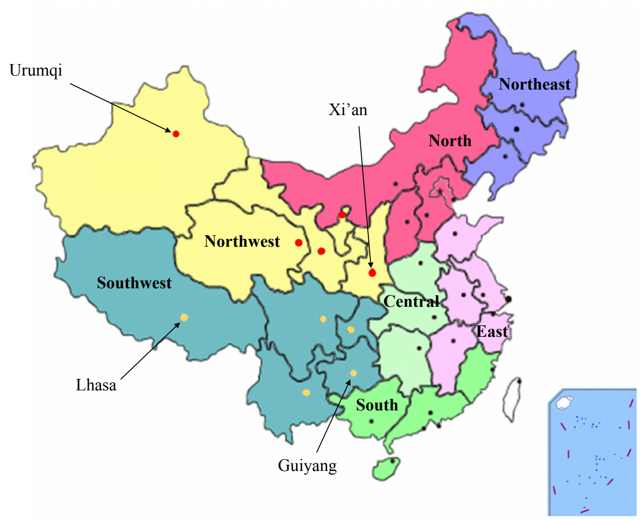


Fig. 12. The seven regions of China. The red dots represent cities in the northwest and the yellow dots represent cities in the southwest.

TABLE IX
THE RESULTS FOR THE FIVE CITIES ARE IN THE SPECIAL TRAINING SET

	Chengdu	Lanzhou	Xi'an	Urumqi	Xining
RMSE	1.88	3.42	3.38	3.71	4.03

to biased estimations.

To verify such hypothesis, two additional experiments are applied specifically to these two regions. For the Southwest, Guiyang and Chongqing are applied for training to test Chengdu. For the Northwest, we plan to augment the network knowledge, so just one testing city is applied, here Lanzhou, Xi'an, Urumqi and Xining are applied as the testing cities separately. Corresponding results are listed in Table IX. The three cities in Southwest are with similar geographical and climatic conditions, thus getting satisfactory results 1.88K. However, for the Northwest, due to the complex geography environment, the testing results are still not satisfactory (3.42K for Lanzhou, 3.38K for Xi'an, 3.71K for Urumqi and 4.03K for Xining). This suggests how to introduce the unique local climatic factors into the network is an interesting point for future research.

2) *Errors from artificial terrain:* It is also worth noting that the artificial flat structures large area often lead to higher errors, such as ports and airports. Fig. 13 (b) and (d), correspond to airports and ports in Xiamen. The dates of LST from left to right in (c) and (e) are captured on March 11, 2018, February 25, 2019, and March 30, 2019, respectively. The climate conditions of these data are similar, however, the temperature represents high dynamic fluctuation. A possible reason could be that, for the flat structures, due to their characteristics of fast heat absorption and dissipation, the temperature tends to change rapidly even at different times of the same day. Such a property makes the urban LST highly correlate to the exact time point of the day when the data is captured, Thus leading to higher prediction error at these areas.

3) *Errors from human behaviour:* Errors related to anthropogenic activities can be observed from our results as well; such errors always occurred in the industrial estates regions.

One typical case is shown in Fig. 14 (a) and (b), in the single city temperature prediction experiment IV-C1, at Hefei city. The region highlighted by the red box in (a) contains

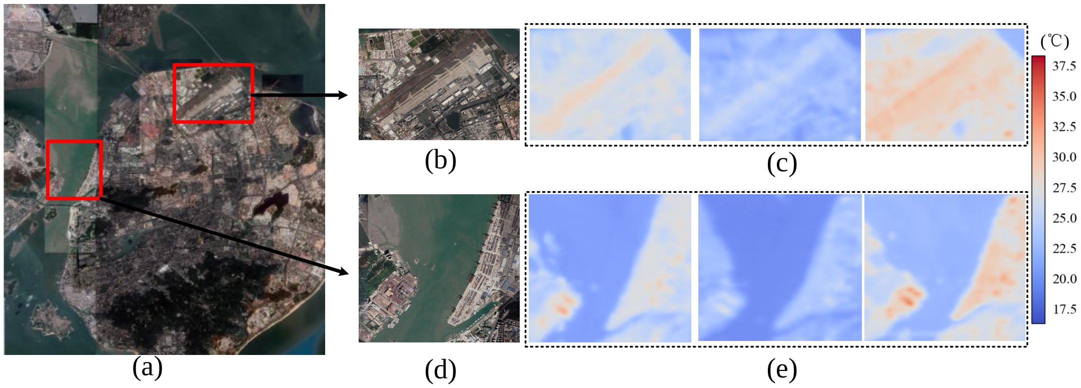


Fig. 13. Illustration of satellite images and LST in Xiamen. (a) Satellite remote sensing images of Xiamen. (b)-(c) Satellite image of Xiamen Gaoqi Airport and the LST data of the corresponding area. (d)-(e) Satellite images of Xiamen Haitian Pier and LST data of the corresponding area.

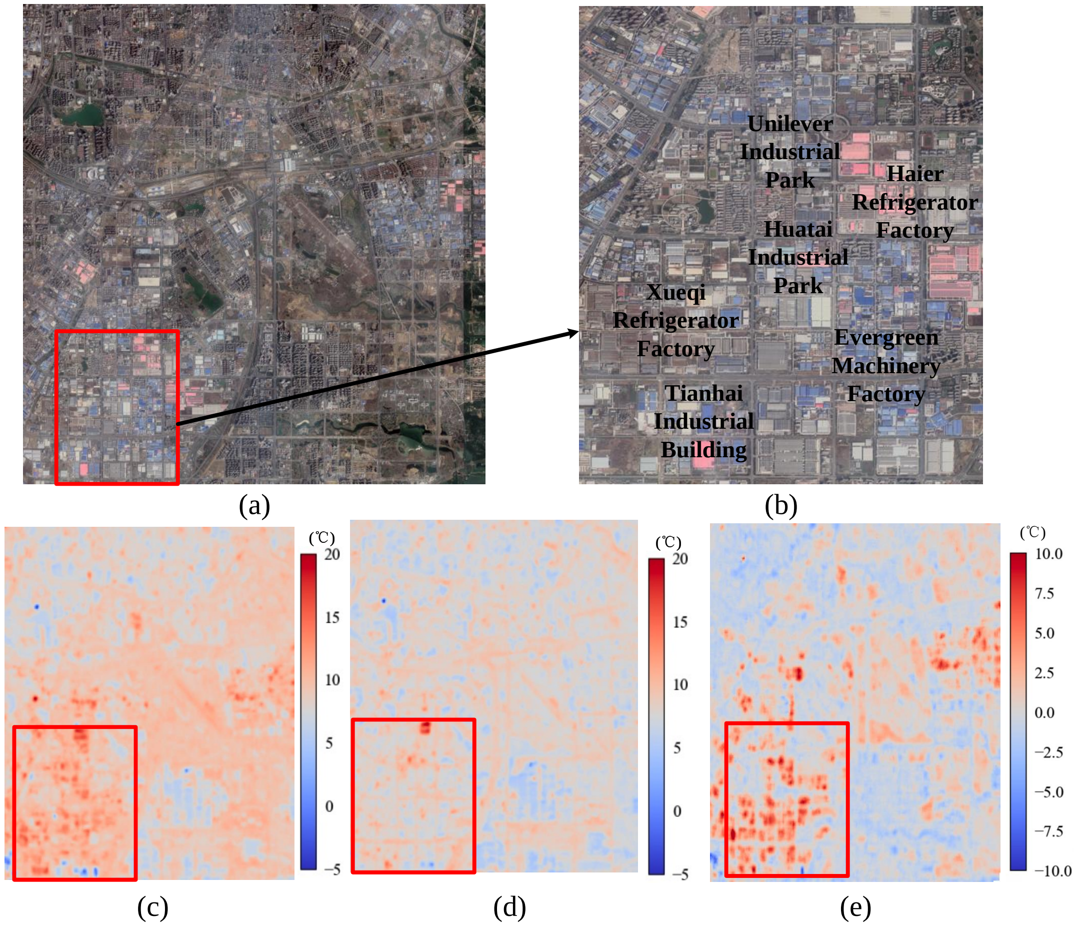


Fig. 14. Illustration of the prediction accuracy analysis of the experimental results in Hefei. (a) Satellite remote sensing images of Hefei. (b) The red boxed area in satellite map (a) contains a large number of factories. (c) LST visualization results for Hefei on Jan 23, 2019. (d) LST visualization results for Hefei on February 05, 2018. (e) Prediction error visualization of Hefei.

a large number of factories and industrial campuses, and the prediction error in this region is shown in Fig. 14 (e), where dense high error points are observed.

The original LST data are visualized for further in-depth error analysis. Figure (c) shows a typical training case, where a region with a local high temperature is observed in the red box area. Figure (d) shows one of two applied testing samples, correspond to the dates of February 5, 2018. The high-temperature region that should have appeared has become less obvious. This is because the date of February 5, 2018 is near to the Chinese New Year holiday, and most of these factories are shut down, leading to higher errors in this region.

Therefore, anthropogenic heat could have large impacts on the urban surface temperatures at such a high resolution, and such a factor should be considered in the future work of intracity temperature prediction.

V. Conclusion

In this paper, we propose a physics informed hierarchical perception network – PIHP-Net – for high-resolution and high-precision urban surface temperature prediction from upper-atmospheric forcing fields and land surface satellite imagery. This network can be used to predict intracity urban LST in future time periods when the future forcing meteorology can be easily obtained from regional climate models. The PIHP-Net employees a designed multi-scale encoder to construct different scales Geographical Semantic Category Fraction Matrix (GSCFM), thus extracting different scales of urban surface structure. Based on this, a hierarchical urban surface perception scheme is proposed via a multi-branch network structure. Such a scheme is able to capture the complex non-linear relationship between the land surface and temperature in different scales. Extensive experimental results show that the PIHP-Net consistently outperforms previous baselines on the high resolution urban LST prediction tasks.

We conducted designed experiments 31 major or provincial capital cities in China to evaluate the performance of the proposed PIHP-Net comprehensively. The estimation errors of the previous static model-based method for 60-by-60 meters grids is about 6 Kelvin, for most cases of our experiments, the errors are less than 2 Kelvin, thus making it possible to estimate or predict *intracity temperature* from multi-spectral satellite imagery. Compared to previous statistical models, the proposed approach has at least more than 30% performance improvement, either for single or multiple city predictions. These results demonstrate the validity and efficacy of our proposed approach.

Major contributions of this paper include the following three aspects:

1) We propose to exploit high-resolution, largely available multi-spectral satellite images to tackle the challenging high-resolution urban temperature estimation problem.
2) We develop, for the first time, a novel physics informed hierarchical perception network, PIHP-Net, for accurate, high-resolution and generalizable urban surface temperature estimation. Benefiting from a proposed novel semantic category histogram and a hierarchical perception scheme, the PIHP-Net can generate accurate urban LST estimation results (at least more than 30% improvement compared to previous statistical model-based works) at a high resolution (60-by-60 meters grids).
3) We build a novel group dataset related to the urban LST. Such datasets cover all the major cities (31 cities were used) in China, which will support comprehensive and large-scale experiments for further research at this field.

Our results also suggest that surface structural properties and human activities greatly influence on the surface temperature at a high resolution, both spatially and temporally. For cities that span a large area, forcing meteorology, as well as urban forms, are the main factors that determine the surface temperature variations.

References

[1] N. B. Grimm, S. H. Faeth, N. E. Golubiewski, C. L. Redman, J. Wu, X. Bai, and J. M. Briggs, "Global change and the ecology of cities," *science*, vol. 319, no. 5864, pp. 756–760, 2008.

[2] C. Mora, B. Dousset, I. R. Caldwell, F. E. Powell, R. C. Geronimo, C. R. Bielecki, C. W. Counsell, B. S. Dietrich, E. T. Johnston, L. V. Louis *et al.*, "Global risk of deadly heat," *Nature climate change*, vol. 7, no. 7, pp. 501–506, 2017.

[3] M. Georgescu, P. E. Morefield, B. G. Bierwagen, and C. P. Weaver, "Urban adaptation can roll back warming of emerging megapolitan regions," *Proceedings of the National Academy of Sciences*, vol. 111, no. 8, pp. 2909–2914, 2014.

[4] E. S. Krayenhoff, M. Moustaoui, A. M. Broadbent, V. Gupta, and M. Georgescu, "Diurnal interaction between urban expansion, climate change and adaptation in us cities," *Nature Climate Change*, vol. 8, no. 12, pp. 1097–1103, 2018.

[5] S. Gaffin, C. Rosenzweig, R. Khanbilvardi, L. Parshall, S. Mahani, H. Glickman, R. Goldberg, R. Blake, R. Slosberg, and D. Hillel, "Variations in new york city's urban heat island strength over time and space," *Theoretical and applied climatology*, vol. 94, no. 1, pp. 1–11, 2008.

[6] H. Li, R. Li, Y. Yang, B. Cao, Z. Bian, T. Hu, Y. Du, L. Sun, and Q. Liu, "Temperature-based and radiance-based validation of the collection 6 MYD11 and MYD21 land surface temperature products over barren surfaces in northwestern china," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1794–1807, 2021.

[7] B. Chun and J.-M. Guldmann, "Impact of greening on the urban heat island: Seasonal variations and mitigation strategies," *Computers, Environment and Urban Systems*, vol. 71, pp. 165–176, 2018.

[8] D. Zhou, J. Xiao, S. Bonafoni, C. Berger, K. Deilami, Y. Zhou, S. Frolking, R. Yao, Z. Qiao, and J. A. Sobrino, "Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives," *Remote Sensing*, vol. 11, no. 1, p. 48, 2019.

[9] X.-M. Zhu, X.-N. Song, P. Leng, D. Guo, and S.-H. Cai, "Impact of atmospheric correction on spatial heterogeneity relations between land surface temperature and biophysical compositions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2680–2697, 2021.

[10] L. Zhao, K. Oleson, E. Bou-Zeid, E. S. Krayenhoff, A. Bray, Q. Zhu, Z. Zheng, C. Chen, and M. Oppenheimer, "Global multi-model projections of local urban climates," *Nature Climate Change*, vol. 11, no. 2, pp. 152–157, 2021.

[11] Z. Zheng, L. Zhao, and K. W. Oleson, "Large model structural uncertainty in global projections of urban heat waves," *Nature Communications*, vol. 12, no. 1, pp. 1–9, 2021.

[12] K. Kashinath, M. Mustafa, A. Albert, J. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh *et al.*, "Physics-informed machine learning: case studies for weather and climate modelling," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200093, 2021.

[13] X. Jia, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, and V. Kumar, "Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles," *ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–26, 2021.

[14] L. Zhao, X. Lee, R. B. Smith, and K. Oleson, "Strong contributions of local background climate to urban heat islands," *Nature*, vol. 511, no. 7508, pp. 216–219, 2014.

[15] E. S. Krayenhoff, A. M. Broadbent, L. Zhao, M. Georgescu, A. Middel, J. A. Voogt, A. Martilli, D. J. Sailor, and E. Erell, "Cooling hot cities: a systematic and critical review of the numerical modelling literature," *Environmental Research Letters*, vol. 16, no. 5, p. 053007, 2021.

[16] W. Yue, J. Xu, W. Tan, and L. Xu, "The relationship between land surface temperature and ndvi with remote sensing: application to shanghai Landsat 7 ETM+ data," *International journal of remote sensing*, vol. 28, no. 15, pp. 3205–3226, 2007.

[17] D. Zhou, J. Xiao, S. Bonafoni, C. Berger, K. Deilami, Y. Zhou, S. Frolking, R. Yao, Z. Qiao, and J. A. Sobrino, "Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives," *Remote Sensing*, vol. 11, no. 1, p. 48, 2018.

[18] T. D. Mushore, O. Mutanga, and J. Odindi, "Estimating urban lst using multiple remotely sensed spectral indices and elevation retrievals," *Sustainable Cities and Society*, vol. 78, p. 103623, 2022.

[19] S. Peng, S. Piao, P. Ciais, P. Friedlingstein, C. Ottle, F.-M. Bréon, H. Nan, L. Zhou, and R. B. Myneni, "Surface urban heat island across 419 global big cities," *Environmental science & technology*, vol. 46, no. 2, pp. 696–703, 2012.

[20] M. L. Imhoff, P. Zhang, R. E. Wolfe, and L. Bounoua, "Remote sensing of the urban heat island effect across biomes in the continental usa," *Remote sensing of environment*, vol. 114, no. 3, pp. 504–513, 2010.

[21] N. Clinton and P. Gong, "Modis detected surface urban heat islands and sinks: Global locations and controls," *Remote Sensing of Environment*, vol. 134, pp. 294–304, 2013.

[22] L. Zhao, K. Oleson, E. Bou-Zeid, E. S. Krayenhoff, A. Bray, Q. Zhu, Z. Zheng, C. Chen, and M. Oppenheimer, "Global multi-model projections of local urban climates," *Nature Climate Change*, vol. 11, no. 2, pp. 152–157, 2021.

[23] H. Kusaka, H. Kondo, Y. Kikegawa, and F. Kimura, "A simple single-layer urban canopy model for atmospheric models: Comparison with multi-layer and slab models," *Boundary-layer meteorology*, vol. 101, no. 3, pp. 329–358, 2001.

[24] M. Georgescu, M. Moustaoui, A. Mahalov, and J. Dudhia, "Summertime climate impacts of projected megapolitan expansion in arizona," *Nature Climate Change*, vol. 3, no. 1, pp. 37–41, 2013.

[25] C. Ru, S.-B. Duan, X.-G. Jiang, Z.-L. Li, Y. Jiang, H. Ren, P. Leng, and M. Gao, "Land surface temperature retrieval from Landsat 8 thermal infrared data over urban areas considering geometry effect: Method and application," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[26] C. Gromke, B. Blocken, W. Janssen, B. Merema, T. van Hooff, and H. Timmermans, "CFD analysis of transpirational cooling by vegetation: Case study for specific meteorological conditions during a heat wave in arnhem, netherlands," *Building and environment*, vol. 83, pp. 11–26, 2015.

[27] A. Middel, N. Chhetri, and R. Quay, "Urban forestry and cool roofs: Assessment of heat mitigation strategies in phoenix residential neighborhoods," *Urban Forestry & Urban Greening*, vol. 14, no. 1, pp. 178–186, 2015.

[28] K. Zakšek and K. Oštir, "Downscaling land surface temperature for urban heat island diurnal cycle analysis," *Remote Sensing of Environment*, vol. 117, pp. 114–124, 2012.

[29] W. P. Kustas, J. M. Norman, M. C. Anderson, and A. N. French, "Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship," *Remote sensing of environment*, vol. 85, no. 4, pp. 429–440, 2003.

[30] S. Bonafoni, "Downscaling of Landsat and MODIS land surface temperature over the heterogeneous urban area of milan," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 2019–2027, 2016.

[31] I. Keramitsoglou, C. T. Kiranoudis, and Q. Weng, "Downscaling geostationary land surface temperature imagery for urban analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1253–1257, 2013.

[32] H. J. Fowler, S. Blenkinsop, and C. Tebaldi, "Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling," *International Journal of Climatology: A Journal of the Royal Meteorological Society*, vol. 27, no. 12, pp. 1547–1578, 2007.

[33] J. Tang, X. Niu, S. Wang, H. Gao, X. Wang, and J. Wu, "Statistical downscaling and dynamical downscaling of regional climate in china: Present climate evaluations and future climate projections," *Journal of Geophysical Research: Atmospheres*, vol. 121, no. 5, pp. 2110–2129, 2016.

[34] S. Spak, T. Holloway, B. Lynn, and R. Goldberg, "A comparison of statistical and dynamical downscaling for surface temperature in north america," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D8, 2007.

[35] W. Li, L. Ni, Z.-L. Li, S.-B. Duan, and H. Wu, "Evaluation of machine learning algorithms in spatial downscaling of modis land surface temperature," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2299–2307, 2019.

[36] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (PGNN): An application in lake temperature modeling," *arXiv preprint arXiv:1710.11431*, 2017.

[37] Y.-G. Ham, J.-H. Kim, and J.-J. Luo, "Deep learning for multi-year enso forecasts," *Nature*, vol. 573, no. 7775, pp. 568–572, 2019.

[38] Z. Zhang, W. Xu, Q. Qin, and Z. Long, "Downscaling solar-induced chlorophyll fluorescence based on convolutional neural network method to monitor agricultural drought," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1012–1028, 2021.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[41] G. He, Z. Zhang, W. Jiao, T. Long, Y. Peng, G. Wang, R. Yin, W. Wang, X. Zhang, H. Liu *et al.*, "Generation of ready to use (RTU) products over china based on Landsat series data," *Big Earth Data*, vol. 2, no. 1, pp. 56–64, 2018.

[42] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.

[43] T. Kshetri, "NDVI, NDBI & NDWI calculation using Landsat 7, 8," *Researchgate. net*, vol. 327971920, 2018.

[44] B. Felbo, A. Mislove, A. Sgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017.

[45] K. Sharma and M. Sharma, "Image fusion based on image decomposition using self-fractional fourier functions," *Signal, image and video processing*, vol. 8, no. 7, pp. 1335–1344, 2014.

[46] A. Averbuch, D. Lazar, and M. Israeli, "Image compression using wavelet transform and multiresolution decomposition," *IEEE Transactions on Image Processing*, vol. 5, no. 1, pp. 4–15, 1996.

[47] N. E. Huang, *Hilbert-Huang transform and its applications*. World Scientific, 2014, vol. 16.

[48] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[49] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Advances in neural information processing systems*, 2017, pp. 971–980.

[50] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.

[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[52] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[53] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

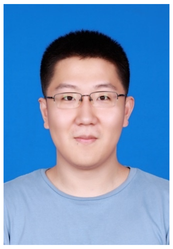
[54] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[55] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

Donghang Wu received the B.S. degree from Fujian Normal University, Fuzhou, China in 2020. He is currently pursuing the M.S. degree with the School of Informatics, Xiamen University, Xiamen, China. His research interests include remote sensing image processing and deep learning.

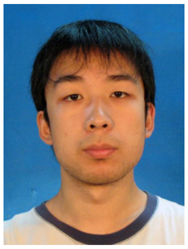
Weiquan Liu received the B.S. and M.S. degrees in applied mathematics from the College of Science, Jimei University, Xiamen, China, in 2016, and recevied the Ph.D. degree in computer science and technology from the School of Informatics, Xiamen University, Xiamen, China, in 2020.

He is currently a Postdoc with the Information and Communication Engineering Postdoctoral Research Station, and the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen, China. His current research interests include remote sensing, computer vision, machine learning, mobile laser scanning point cloud data processing, and augmented reality.

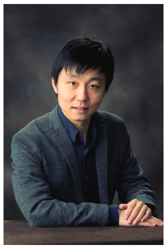


Bowen Fang received his B.S. and M.S. in Environmental Science from Peking University and Yale School of the Environment. He is currently a PhD student at Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign. His research uses earth system model and remote sensing to study the interactive impact of urbanization and climate change. He is also interested in urban heat island mitigation and other engineering solutions to sustainable urban development.

Linwei Chen received the B.S. degree from Fujian Normal University, Fuzhou, China in 2019. He is currently pursuing the M.S. degree with the School of Informatics, Xiamen University, Xiamen, China. His research interests include deep learning and its applications on 3D point cloud processing.



Yu Zang is currently a Research Associate professor at the School of Informatics, Xiamen University, China. He received his B.S. and Ph.D. degree in Xi'an Jiaotong University in 2008 and 2014. His main researches include remote sensing image processing, computer vision&graphics and mobile LiDAR data analysis.



Lei Zhao is currently an Assistant Professor in the Department of Civil and Environmental Engineering and Assistant Professor affiliated with the National Center of Supercomputing Applications at the University of Illinois at Urbana-Champaign. He received his B.S. degree in Atmospheric Physics from Nanjing University in 2009, and Ph.D. degree from Yale University in 2015. Before joining the University of Illinois at Urbana-Champaign, Lei Zhao finished his postdoctoral training at Princeton University. His research interests include multi-scale climate modeling, remote sensing, urban climate, environmental fluid mechanics ann turbulence, machine learning and statistical modeling. He has published more than 18 peer-reviewed papers in worldwide top-ranked journals including Nature, Nature Climate Change, Nature Geoscience, Nature Communications as first author and/or corresponding author.

Shenlong Wang is currently an Assistant Professor at the UIUC Department of Computer Science, affiliated with Computer Vision @ UIUC and Illinois Robotics Group. Before joining UIUC, I was a visiting scholar at the Intel Intelligent Systems Lab, working with Vladlen Koltun. I finished my PhD at Department of Computer Science, University of Toronto and worked as a research scientist at Uber ATG, working with Raquel Urtasun. His research interests include 3D computer vision and robotics, specifically on topics such as 3D perception and modeling, photorealistic content creation, localization and mapping, robot simulation and self-driving.

Cheng Wang (M'07-SM'16) received the Ph.D. degree in signal and information processing from the National University of Defense Technology, Changsha, China, in 2002.

He is currently a Professor with the School of Informatics, and the Executive Director with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen, China. He has coauthored more than 150 papers in referred journals and top conferences including IEEE Transactions on Geoscience and Remote Sensing, PR, IEEE Transactions on Intelligent Transportation Systems, IEEE Conference on Computer Vision and Pattern Recognition, Association for the Advancement of Artificial Intelligence (AAAI), and International Society for Photogrammetry and Remote Sensing (ISPRS) Journal of Photogrammetry and Remote Sensing. His current research interests include point cloud analysis, multisensor fusion, mobile mapping, and geospatial big data.

Prof. Wang is a Fellow of the Institution of Engineering and Technology. He is also the Chair of the Working Group I/6 on Multi-Sensor Integration and Fusion of the International Society of Remote Sensing.

José Marcato Junior (Member, IEEE) received his Ph.D. degree in 2014 from São Paulo State University, Brazil. He is a Professor with the Faculty of Engineering, Architecture and Urbanism and Geography at the Federal University of Mato Grosso do Sul (UFMS) in Brazil. He is the head of the Geomatics Laboratory at UFMS and has experience in coordinating international projects. With expertise in photogrammetry and remote sensing, his current research program is focused on UAV and close-range photogrammetry, and deep learning in environmental and precision agriculture applications. He has published over 75 peer-reviewed journal papers, and has experience as associate editor of two scientific journals and guest editor of several special issues in refereed journals of remote sensing. He becomes in 2020 a distinguished researcher by the Brazilian CNPq (Brazilian National Council for Scientific and Technological Development).

Jonathan Li received the Ph.D. degree in geomatics ngineering from the University of CapeTown, South Africa, in 2000. He is currently a Professor of geomatics and systems design engineer in gat the University of Waterloo, Canada. He is also Founding Member of the Waterloo Artificial Intelligence Institute. His research interests in clude AI-based informatione xtraction from mobile LiDAR point clouds and Earth observation images. He has coauthored more than 450 publications, more than 260 of which were published in refereed journals, including IEEE Transactionson Geoscience and Remote Sensing, IEEE Transactionson Intelligent Transportation Systems, ISPRS Journal of Phorogrammetry and Remote Sening, and Remote Sensing of Environment. He is currently the Editor in Chief of the International Journal of Applied Earth Observation and Geoinformation, Associate Editor of the IEEE Transactionson Intelligent Transportation Systems, IEEE Transactions on Geoscience and Remote Sensing, and Canadian Journal of Remote Sensing. He was a recipient of the ISPRS Samuel Gamble Award in 2020