DEEPURBANMODELLER (DUM): A PROCESS-INFORMED NEURAL ARCHITECTURE FOR HIGH-PRECISION URBAN SURFACE TEMPERATURE PREDICTION

Yu Zang[1], Guanjie Huang[1], Weiquan Liu,[*] Linwei Chen[1], Donghang Wu[1], Cheng Wang[1], Jonathan Li[2]

[1] Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen, China
[2] GeoSTARS Laboratory, Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada

KEY WORDS: Deep urban downscale, Physics informed , Neural network, Point cloud, Land surface temperature.

ABSTRACT:

High-resolution downscaling of surface climate metrics like urban surface temperature, is a crucial and ongoing research challenge in urban climatology and environmental studies. In this study we propose a groundbreaking Physics-Inspired Neural Architecture for Modeling (PINAM) called DeepUrbanModeller(DUM), designed specifically for urban microclimate temperature estimation. DeepUrbanModeller(DUM) harnesses process-based modelling and satellite remote sensing, and draws upon high-accuracy 3D point clouds to deliver precise estimations of urban Land Surface Temperature (LST) at ultra-high resolutions. By incorporating high-accuracy land surface geometric data sourced from 3D point clouds and guided by the principles of atmospheric physics linked to surface temperature, DeepUrbanModeller(DUM) creates a data-driven framework, informed by physical laws, to accurately model high-resolution temperature distributions a task challenging for numerical simulations or conventional machine learning. The DeepUrbanModeller(DUM) design integrates two key components: Global Physical Feature Interpretation (GPFI) and Local Urban Surface Insight (LUSI). The GPFI captures broader urban physical parameters, ensuring the estimates comply with relevant physical laws. The LUSI enhances estimation performance at high-resolution levels by utilizing a newly proposed Urban Detail Orientation Index (UDOI) derived from 3D point clouds. Experimental results demonstrate the DeepUrbanModeller(DUM)'s superior capability in estimating urban LST on a detailed 30-by-30 meter grid, achieving an estimation error of less than 0.2 Kelvin compared to satellite measurements, a performance surpassing traditional methodologies.

1. INTRODUCTION

Cities, being densely populated and infrastructure-rich, are heavily impacted by climate change (Grimm et al., 2008), (Daw et al., 2017)) . Effective urban planning requires high-resolution, accurate climate predictions, with Land Surface Temperature (LST) being critical due to its effects on public health, energy management, infrastructure safety, and resilience against extreme weather events (Georgescu et al., 2014).The temperature at the surface of urban lands, also known as Land Surface Temperature (LST), is a key climatic factor that garners significant public attention. Its relevance is directly linked to several important issues such as alterations in urban climate, public health implications, strategies for urban energy management, safeguarding infrastructure, and the resilience of systems to extreme weather events. For instance, extremely high temperatures in urban areas can trigger a notable surge in instances of death and illness among humans.(Anderson and Bell, 2011)(Patz et al., 2005)(Huang et al., 2011)energy demand and power grid failure(Isaac and Van Vuuren, 2009).

In this study we will introduces the High Resolution Urban Forecaster (HRUF), a physics-inspired neural architecture, for improving the granularity of urban LST predictions. The HRUF leverages procedural modeling and satellite-based remote sensing, using detailed 3D point clouds for highly accurate LST forecasts. By considering atmospheric stimuli and precision land surface topographic data, HRUF maintains crucial physics (Zhao et al., 2014).The HRUF model incorporates both the detailed physics of a typical dynamic downscaling model and the super high-resolution urban surface characteristics, leading to high precision and wide spatial adaptability. Compared with es-

tablished downscaling strategies, HRUF has achieved improved spatial resolution (from 1000m to 30m) and reduced estimation divergence (below 0.2 Kelvin), surpassing current high-grade downscaling procedures.

2. RATIONALE AND PROPOSED FRAMEWORK

A. Rationale

The central understanding behind a proposed algorithm which is related to key physical processes that determine urban surface temperature. These processes, involving interactions between urban land and atmosphere, include incoming and reflected solar radiation, longwave radiation exchanges, heat transfer, and evapotranspiration from permeable surfaces, along with the heat stored in buildings.

These processes are significantly influenced by atmospheric pressure conditions and surface properties. Dynamic models, based on these processes, utilize atmospheric pressure variables and urban surface data to solve physical equations for Land Surface Temperature (LST), albeit at a broad spatial resolution. High-resolution modeling is hindered by the complexity of the urban surface and the prohibitive computational load required for large domain applications, rendering such models almost impractical. Based on the considerations described above, the rationale of this study can be summarized as:

- For fundamental landscapes like metropolitan regions, the inherent physical procedures can be extraordinarily intricate. This complexity makes it nearly unfeasible to accurately estimate the LST by deciphering the comprehensive physics.

---

* Corresponding author, Email: wqliu@xmu.edu.cn

(a) Zhang Zhou Harbor satellite imagery

(b) Land surface temperature

(c) Our estimation
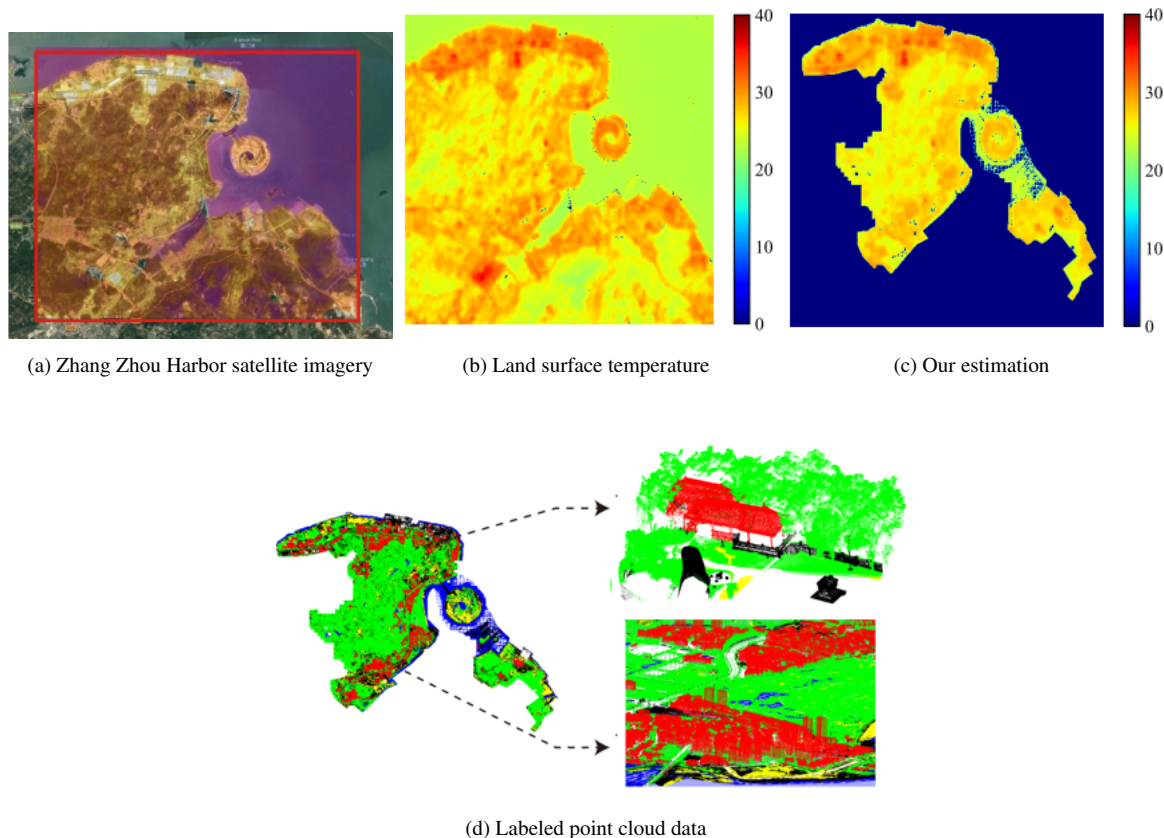

(d) Labeled point cloud data

Figure 1. Fig. 1. Overview of the datasets and the results. (a) Testing area of Zhang Zhou Harbor. (b) Visualized map of the land surface temperature (LST) captured by Landsat satellite. (c) Our estimation result. (d) Visualized map of the labeled 3D point cloud.

- This study is aim to delve into the intrinsic associations between the LST and all pivotal components involved in the dynamic physical procedures. This is achieved by employing our specially designed physics-aware deep neural network - DeepUrbanModeller(DUM).

These rationale further implies the necessary datasets (see Section 3-A).

B. Framework Overview

Land Surface Temperature (LST), which is governed by various biophysical processes related to the urban surface energy balance, including both shortwave and longwave surface radiation balance, the turbulent transport of sensible heat, and surface evapotranspiration. These factors are contingent on atmospheric factors such as solar and atmospheric radiation, air temperature, wind, air pressure, and humidity, and local urban surface properties such as greenery, surface roughness, building heights, and the layout of buildings and streets. Broad-scale urban surface climate, like average citywide temperature, is primarily determined by atmospheric forcings, while local-scale variations are mostly influence urban surface features. The DeepUrbanModeller (DUM) network is designed to reflect these physical principles, consisting of two branches: the Global Physical Feature Interpretation (GPFI) branch and the Local Urban Surface Insight (LUSI) branch, as illustrated in Figure 2.

The Global Physical Feature Interpretation (GPFI) branch is designed to establish correlations between meteorological and climatological factors, including synoptic conditions, climate shifts, and seasonal variations. However, it falls short in providing precise high-resolution temperature predictions. To address this, the Local Urban Surface Insight (LUSI) branch refines the results by capturing high-resolution variability. A notable advancement in the LUSI branch is the inclusion of high-definition urban 3D point cloud data. Traditional methods of downscaling Land Surface Temperature (LST) have predominantly relied on 2D surface property data like the Normalized Difference Vegetation Index (NDVI). Yet, research has shown that the 3D geometry of a surface, such as surface roughness, significantly affects urban surface temperature.

C. Global Physical Feature Interpretation Branch (GPFI)

The GPFI (Global Physical Feature Interpretation) branch integrates key atmospheric variables from urban climate models into a deep neural network,(Oleson et al., 2008). These variables, detailed in Table 1, are processed using a multi-layer perceptron (MLP) sourced from the MERRA-2 reanalysis dataset (Gelaro et al., 2017). The MLP, consisting of four layers and utilizing the ReLU function, effectively emulates the model dynamics and aids in solving physical equations. The GPFI's role is to guide initial temperature predictions to align with atmospheric conditions, with further refinement by the Local Surface Interpretation branch.

D. Local Urban Surface Insight Branch (LUSI)

The Urban Detail Orientation Index (UDOI) encapsulates local 3D urban geometry in the neural network, eschewing full 3D point cloud input (Qi et al., 2017), (Thomas et al., 2019), (Xu et al., 2020). This approach addresses system generalization challenges arising from urban surface complexity.UDOI comprises a surface property index and a local geometry index. The surface property index is defined by the proportion of
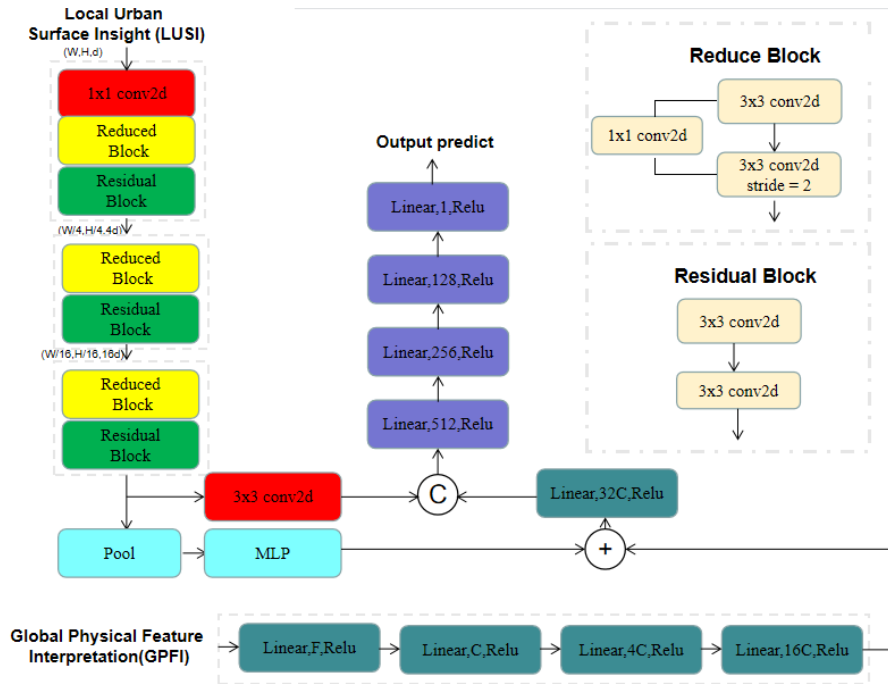
Figure 2. Architecture of the proposed DeepUrbanModeller system.

specific structure categories (water, buildings, vegetation, soil, and roads/pavements) within a region, following established semantic labeling methods (Hackel et al., 2017) (Fig. 3(a)).The local geometry index derives from the average cell height, multiplied by the urban building index. It encapsulates spatial layout, surface roughness, and building verticality, which are crucial for understanding local atmospheric turbulence over urban surfaces.

Specifically, the whole UDOI is defined by the following equation:

$$Ides(l) = \frac{C(S_l)}{C(S)}, \quad S \in T.\rho \tag{1}$$

$$Hdes(l_b) = Ides(l_b) \cdot avg_z(S_{l_b}) \tag{2}$$

where S represents point cloud set centered at a certain cell, $C(\cdot)$ denotes the number of points in S, S_l denotes the number of points in S with certain category l, l_b denote the label of building, and $avg_z(\cdot)$ calculates the average height of all points in S.

In Figure 3(b) and (c), we see a sample construction of the UDOI. For every 30-by-30 meter grid, we generate a standard m×m×d matrix, where 'd' symbolizes the feature vector dimension at a given cell. This matrix defines the Urban Detail Orientation Index (UDOI).

The Local Urban Surface Insight (LUSI) branch leverages the Urban Detail Orientation Index (UDOI) to extract local surface characteristics, including high-resolution variability of urban temperature. It comprises a deep residual network with four stages, each having two residual blocks. Convolution layers, strides, skip connections, and batch normalization are judiciously employed to ensure feature preservation.

UDOI matrix expansion is achieved by encompassing a k-meter range around the central grid. This enhancement aids

in error reduction by incorporating context information. The processed features and 3D structure information are encoded into a $1 \times 32C$ size latent vector.

Finally, outputs from LUSI and GPFI branches are combined into a latent feature vector of size $32C+16d$, encapsulating atmospheric forcing factors, high-resolution urban surface features, and 3D geometric structure information. This is directed into a regression branch with three fully-connected layers.

E. Loss Functions

Following the previous work (Klambauer et al., 2017), we employ the mean squared error and the L_2 normalization of the network weights to measure the loss. The overall loss can be written as:

$$\underset{W,b}{\arg\min} : L(Y, \hat{Y}) + \lambda R(W) \tag{3}$$

$$L(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4}$$

$$R(W) = ||W||^2 \tag{5}$$

Here, Y and $\hat{Y}$ represent the ground truth set and predicted results set, W and b are the combined coefficient of weights and bias terms, $y_i \in Y$, $\hat{y}_i \in \hat{Y}$, $i = 1, ..., N$, N is the data size of Y, $\hat{Y}$, λ is the weight of regularization term.

3. EXPERIMENT RESULTS

A. Datasets Description

For this research, we focused on Zhang Zhou Harbor, a modestly sized city in China, as our test site. Fig. 1(a) reveals that this region spans an area of 150.56 square kilometers and hosts a diverse set of landscapes, including urban zones, mountainous regions, and bodies of water (indicated within the

Table 1. ATMOSPHERIC AND LOCATION DATA.

Feature information	Unit
1 Surface absorbed longwave radiation	$W \cdot m^{-2}$
2 Incident shortwave land	$W \cdot m^{-2}$
3 Total precipitation over land	$kg \cdot m^{-2} \cdot s^{-1}$
4 Surface pressure	Pa
5 Air temperature at the reference height	K
6 Eastward wind	m^{-1}
7 Northward wind	m^{-1}
8 Specific humidity	$kg \cdot kg^{-1}$
9 Longitude	km
10 Latitude	km

red box). To drive our DeepUrbanModeller (DUM) model, we assembled a collection of datasets sourced from multiple providers within the studied region, as elaborated below.

- LST. Our primary dataset encompasses land surface temperature data computed from NASA's Landsat satellite imagery. Landsat offers an extensive record of Earth's terrestrial areas since 1973. The data incorporates various band types (Blue, Red, Green, Near Infrared, etc.), and specific bands are utilized to estimate LST through established algorithms [38], [39]. This dataset can be publicly accessed from the USGS2 website. For this study, we collected data spanning from July 2013 to July 2020, with a 16-day interval. The dataset's spatial resolution stands at a 30-by-30 meter grid. Fig. 1(b) provides a visualization, where each pixel signifies the LST of a 30-by-30 meter space.

- NDVI. The second dataset includes the Normalized Difference Vegetation Index (NDVI) data, also procured from NASA's Landsat satellite. NDVI values range from -1 to 1, with larger values indicating denser vegetation. We curated the data to match the temporal interval and location of the LST dataset. The resolution of this dataset is also a 30-by-30 meter grid. It can be accessed from the USGS website or NASA's MODIS3.

- Atmospheric forcing. The third dataset pertains to atmospheric forcing data, sourced from NASA's MERRA-2 reanalysis data system [32]. This data outlines the comprehensive characteristics of specific areas, owing to each region's unique atmospheric features. The main components of the atmospheric forcing data are outlined in Table 1. Publicly available on the NASA MERRA-2 website4, the resolution of this data is 0.5° latitude x 0.625° longitude.

- Land surface 3D structure. The final dataset encompasses the 3D point cloud data of the entire Zhang Zhou Harbor region, delineating the accurate 3D structure of the area. Constructed using the DaJiang Inspire-1 UAV and the RIEGL VMX-450 mobile laser scanning system, which can produce 1.1 million range measurements per second and acquire nearly 100GB of point cloud data in an hour, this data includes varied scenes like urban areas, towns, and villages. The point cloud data is manually labeled into eight main categories: water (blue), buildings (red), vegetation (green), soil (yellow), roads (gray), pavement (white), vehicles (purple), and others (black).

Finally, we align all data, let each cell (a 30-by-30 meter area) data can be described by a set of attributes denoted by 4-tuple:

$$T = \{t | t = (\tau, \eta, \alpha, \rho)\} \qquad (6)$$

where τ, η, α, ρ are the value of LST, NDVI, atmosphere features and the set of point cloud. We have open-sourced all of the datasets, they can be downloaded from FTP server.
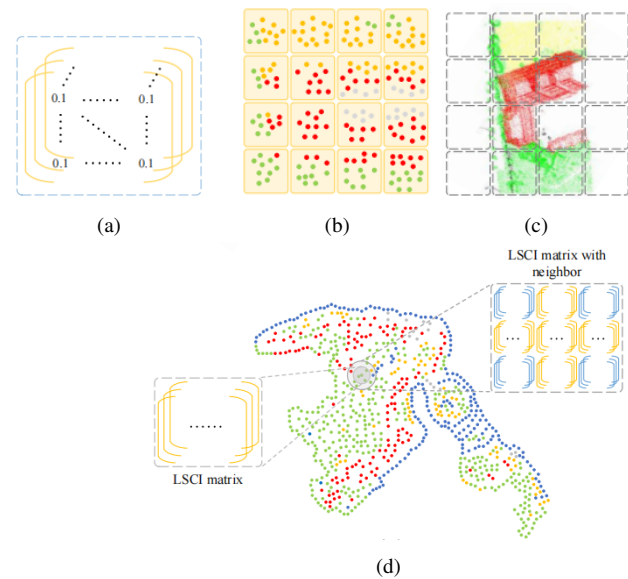

Figure 3. The illustration how to generate the LSCI. (a) An example of the labeled point cloud for a 30-by-30 meter grid. (b) Visualized result of different categories in a certain cell. (c) Example of the LSCI matrix with size m × m × d. (d) Illustration of the aggregation for a grid.

B. Evaluations of the proposed DUM system

(1) Performance.: Our method's effectiveness was evaluated using ten data sets spanning various seasons throughout a year. We utilized the root mean squared error (RMSE) as a performance metric. The performance of our proposed approach is gauged in Kelvin units, with a 70% training data allocation. The associated results are presented in Table 2. These findings indicate that the average error across the ten data sets is approximately 0.11K (as shown in the last column), signifying our method's ability to deliver consistently precise outcomes regardless of seasonal variations. The related visualizations are displayed in Fig. 4. The first line depicts the ground truth (with a color gradient from deep blue to deep red representing temperatures from 0 Celsius to 40 Celsius), while the second line presents a map of estimation errors, highlighting different error levels (ranging from 0K to 1.8K) with various colors.

Quarter-1	Quarter-2	Quarter-3	Quarter-4	Avg. Error(K)
0.102	0.122	0.107	0.109	0.116

Table 2. THE RESULTS OF OUR APPROACH IN DIFFERENT SEASONS.

(2) Effectiveness of the UDOI: In this section, to confirm the effectiveness of the proposed Urban Detail Orientation Index (UDOI), we employ a point cloud-oriented network, PointNet, as a benchmark. We modify the Local Urban Surface Insight (LUSI) module using the framework of PosPool, which incorporates point cloud into the deep residual network. The introduction of point cloud data is shown to reduce the estimation error, though this enhancement is marginal. The corresponding
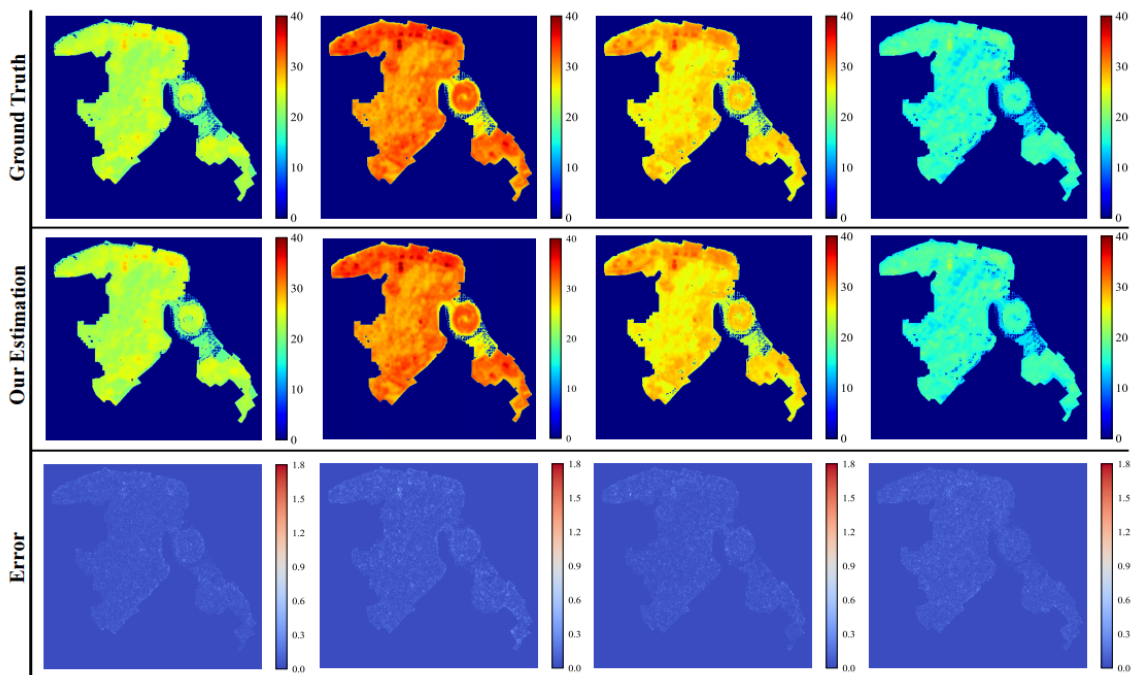
Figure 4. The visualized results of the ground truth, our approach and the error map for quarter 1 to 4

Table 3. HOW THE UDOI AFFECTS THE RESULTS.

Without LUSI branch	LUSI branch based on point cloud	LUSI branch based on the UDOI
Train/Test error 1.103/1.168 K	Train/Test error 0.701/1.002 K	Train/Test error 0.112/0.122 K

statistics are outlined in Table 3. The direct inclusion of point cloud data into the network improves the result by 0.166K, but this method results in significant overfitting. In contrast, substituting raw 3D point cloud data with the proposed UDOI significantly enhances the performance of the network.

C. Comparison with Traditional Method

To evaluate our approach, we have compared it with several traditional Land Surface Temperature (LST) downscaling methods such as linear regression, KNN regression, and random forest regression, all of which were implemented based on Scikit-learn. These traditional methods are known to have high computational cost and, while they offer a reasonable accuracy of about 1K, they cannot achieve high spatial resolution over city-scale coverage. Hence, we shifted our focus on comparing with these statistical downscaling techniques.

We selected the same ten datasets that cover different seasons of the year 2017 to assess these various methods. For a fair comparison, we have incorporated our proposed Urban Detail Orientation Index (UDOI) into all the methods to encapsulate the local geometric information. In each method, we reshaped each dimension of the $m \times m \times d$ matrix to a $1 \times d$ vector and imported this vector into various regression methods.

For the linear regression model, the average error was over 1K due to the lack of clear linear relationships between the variables. When it comes to KNN regression and random forest regression, we manually adjusted the hyper-parameters to best fit the ten datasets. Specifically, the number of neighbors in the KNN regression was set to 4, and in the random forests regression, the maximum tree depth and the number of trees were set to 30 and 150, respectively.

The results show that traditional machine learning methods perform much better than the linear regression model, with errors lower than 1K. However, these methods may suffer from generalizability issues when applied to multiple cities on a large scale, due to their hyper-parameter settings.

As for our proposed DeepUrbanModeller (DUM) network, under 50% training samples, the average estimation error is about 0.13K, which is significantly lower than the error magnitude of the traditional statistical downscaling methods tested above. We plan to conduct more comprehensive testing across various cities in the future.

Table 4. AVERAGE ERROR FOR 10 PIECES OF DATA FOR DIFFERENT SEASONS IN YEAR 2017.

Method	Feb.	Jun.	Oct.	Dec.	Avg. Error(K)
Linear Reg	1.634	1.35	1.586	1.533	1.304
KNN Reg	1.099	0.828	1.041	1.031	0.861
RF Reg	0.541	0.452	0.539	0.46	0.426
DUM	0.156	0.151	0.144	0.149	0.131

4. DISCUSSION

- **Contribution.** From the Experiment section 3, we find that the proposed DUM system performs considerably well for the testing area – Zhang Zhou Harbor. Benefiting from the incorporation of the 3D point cloud, average errors about 0.15K are observed with 50% of data as training samples. Major contributions of this work include:

  1. It provides a first-of-its-kind solution of surface temperature downscaling over highly-complex urban areas

by implementing a PINAM-based architecture to incorporate both process-based insights and data-driven information.

2. Future extension of this work to larger-scale domains (such as regional, national and global scales) and to include more predicted physical quantities (such as surface solar radiation, turbulence, surface wind speed, etc.) would bring new inspirations to the global climate change, energy flow, and other fields.

- **Limitation.** One limitation of the proposed work lies in the range of the testing area. Due to the data availability (labeling high precision 3D point cloud data is labor intensive and thus rather limited for larger-scale experiments), the experiment of the DUM system at the current stage focuses on a single city in China to validate the algorithm.

- **Future Work.** Firstly, our work of expanding the DUM system to cover more cities over a much larger domain is underway.Secondly,we prepare to incorporate more urban variables, as well as design improved feature extraction methods for urban 3D structure point clouds.

5. CONCLUSION

In this paper, we propose a PINAM-based framework, the DUM network, for high-resolution, high-precision urban surface temperature downscaling. The DUM network leverages the Global Physical Feature Interpretation (GPFI) branch to capture broader-scale influences by the atmospheric forcings. Furthermore, the Local Urban Surface Insight (LUSI) branch extracts the high-precision land surface geometry information by employing a proposed Urban Detail Orientation Index (UDOI). With both modules, the DUM network achieves high-accuracy temperature prediction with the estimated error of less than 0.2K.

The DUM network combines process-based modeling and deep learning approaches to provide ultra-high resolution urban LST predictions in a computationally efficient manner. This network can be adopted in other urban surface climate prediction applications that otherwise would require either computationally-expensive (and maybe unattainable) dynamic downscaling or less-accurate traditional statistical methods.

For future work,we prepare to incorporate more urban variables, as well as design improved feature extraction methods for urban 3D structure point clouds,and expand the DUM system to cover more cities

6. ACKNOWLEDGEMENTS

REFERENCES

Anderson, G. B., Bell, M. L., 2011. Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 US communities. *Environmental health perspectives*, 119(2), 210–218.

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

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R. et al., 2017. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, 30(14), 5419–5454.

Georgescu, M., Morefield, P. E., Bierwagen, B. G., Weaver, C. P., 2014. Urban adaptation can roll back warming of emerging megapolitan regions. *Proceedings of the National Academy of Sciences*, 111(8), 2909–2914.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., Briggs, J. M., 2008. Global change and the ecology of cities. *science*, 319(5864), 756–760.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., Pollefeys, M., 2017. Semantic3d.net: A new large-scale point cloud classification benchmark. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1-W1, 91–98.

Huang, C., Barnett, A. G., Wang, X., Vaneckova, P., FitzGerald, G., Tong, S., 2011. Projecting future heat-related mortality under climate change scenarios: a systematic review. *Environmental health perspectives*, 119(12), 1681–1690.

Isaac, M., Van Vuuren, D. P., 2009. Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. *Energy policy*, 37(2), 507–521.

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

Oleson, K. W., Bonan, G. B., Feddema, J., Vertenstein, M., Grimmond, S., 2008. An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities. *Journal of Applied Meteorology and Climatology*, 47(4), 1038–1060.

Patz, J. A., Campbell-Lendrum, D., Holloway, T., Foley, J. A., 2005. Impact of regional climate change on human health. *Nature*, 438(7066), 310–317.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Xu, Q., Sun, X., Wu, C., Wang, P., Neumann, U., 2020. Gridgcn for fast and scalable point cloud learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhao, L., Lee, X., Smith, R. B., Oleson, K., 2014. Strong contributions of local background climate to urban heat islands. *Nature*, 511(7508), 216–219.