Deep learning-based urban morphology for city-scale environmental modeling

Pratiman Patel [ID][a,b], Rajesh Kalyanam[c], Liu He[a], Daniel Aliaga[a] and Dev Niyogi[d,e,*]

[a]Department of Computer Sciences, Purdue University, 305 N University St, West Lafayette, 47907 IN, USA
[b]Interdisciplinary Programme in Climate Studies, Indian Institute of Technology Bombay, Powai, Mumbai, 400076 Maharashtra, India
[c]Research Computing, Purdue University, 155 S Grant St, West Lafayette, 47907 IN, USA
[d]Department of Geological Sciences, Jackson School of Geosciences, The University of Texas at Austin, 2305 Speedway Stop C1160, Austin, 78712-1692 TX, USA
[e]Department of Civil, Architectural, and Environmental Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, TX, USA
*To whom correspondence should be addressed: Email: happy1@utexas.edu
Edited By: Jiahua Zhang

Abstract

Herein, we introduce a novel methodology to generate urban morphometric parameters that takes advantage of deep neural networks and inverse modeling. We take the example of Chicago, USA, where the Urban Canopy Parameters (UCPs) available from the National Urban Database and Access Portal Tool (NUDAPT) are used as input to the Weather Research and Forecasting (WRF) model. Next, the WRF simulations are carried out with Local Climate Zones (LCZs) as part of the World Urban Data Analysis and Portal Tools (WUDAPT) approach. Lastly, a third novel simulation, Digital Synthetic City (DSC), was undertaken where urban morphometry was generated using deep neural networks and inverse modeling, following which UCPs are re-calculated for the LCZs. The three experiments (NUDAPT, WUDAPT, and DSC) were compared against Mesowest observation stations. The results suggest that the introduction of LCZs improves the overall model simulation of urban air temperature. The DSC simulations yielded equal to or better results than the WUDAPT simulation. Furthermore, the change in the UCPs led to a notable difference in the simulated temperature gradients and wind speed within the urban region and the local convergence/divergence zones. These results provide the first successful implementation of the digital urban visualization dataset within an NWP system. This development now can lead the way for a more scalable and widespread ability to perform more accurate urban meteorological modeling and forecasting, especially in developing cities. Additionally, city planners will be able to generate synthetic cities and study their actual impact on the environment.

Keywords: deep neural network, weather research and forecasting model, urban climate, urban boundary layer, WUDAPT, urban canopy parameters

---

Significance Statement

Urban meteorological simulations in Numerical Weather Prediction or Earth System Models are sensitive to the representation of urban areas and the input of high-resolution datasets. Paradoxically, these datasets are hard to obtain and often require specialized measurements—which provide snapshots of a small domain. Additionally, the representation of urban area necessitates the translation of the urban extent to parameters that interact with the model formulations. Hence, this study provides a method to generate urban parameters using global datasets that can lead the way for a scalable ability to perform accurate urban meteorological modeling and forecasting, especially for data scarce regions.

Introduction

Urban centers are economic hubs of the world that contribute about 60% of the global gross domestic product while accommodating more than half of the world's population (1). As urbanization intensifies, cities experience extreme weather conditions such as compound flooding (2), and heatwaves (3). Additionally, urbanization has led to changes in weather conditions such as rainfall (4), urban heat island (5), and air pollution (6). Different

Urban Canopy Models (UCMs) have been developed to incorporate and study urban interaction with the environment. Examples include single layer (7), multi-layer (8, 9), town energy balance (10), and community land model urban parameterization (11).

UCMs utilize urban morphometric details and provide a more realistic urban representation aiding the weather/climate model's performance in simulating urban environments (12). The single-layer UCM incorporates detailed physics of the radiation representation, turbulent transportation, and assumes infinitely

---

long streets for urban geometry representation. Multi-layer UCMs are more sophisticated compared to the single-layer UCM in representing buildings and street layout. The energetics and internal wind as well as thermal effects are represented to provide a more explicit linkage with the urban canopy boundary layer, and coupling with the atmospheric surface and boundary layer (13, 14). The urban models can be run in offline mode to conduct energy balance studies (15, 16) or can be coupled with numerical weather prediction (NWP) or regional climate framework such as, Regional Atmospheric Modeling System (RAMS, (17, 18)), COSMO-CLM model (19), and the Weather Research and Forecasting model (WRF; (7)).

This study considers the WRF urban modeling framework. The WRF model is a mesoscale, non-hydrostatic, compressible, and one of the most widely used NWP model worldwide for operational and research purposes (20). The Noah land surface model used in the WRF model can be coupled with different urban parameterization schemes driven by approximately 30 input parameters for representing urban canopy (21). These Urban Canopy Parameters (UCPs) represent thermal, and geometric properties (including the building internal temperature), which play a role in urban boundary layer (22), precipitation (23), and urban heat island (24) simulations. In an ideal case scenario, the UCPs can be specified for each grid cell where such dataset is specifically generated and made available. If not, as on the default case, the UCPs are specified in the form of a look-up table in the standard WRF framework, for three broad urban classes (low-intensity residential, high-intensity residential, and commercial), which notably underestimate urban complex morphology.

Due to the lack of detailed UCPs, the urban NWP modeling community has been recently shifting towards the use of Local Climate Zones (LCZs) and the World Urban Database and Access Portal Tools (WUDAPT) initiative, which divide the urban surface into 10 different classes (commonly termed as LCZs) based on building height, the density of the buildings, vegetation fraction, and material thermal properties (25, 26). LCZs are being used in the numerous studies focusing on temperature (27), rainfall (12), and other environmental variables (28).

However, developing UCPs for these various urban classes is challenging. This is because mapping urban areas at street level for urban morphology is expensive, requires coordination and approvals from different agencies. Even when such data are available, integrating such data within urban models can be challenging. Some of the examples of a large, community efforts through the US EPA, wherein UCPs are available for 44 US cities comprising the National Urban Database and Access Portal Tool (NUDAPT; (29)). There are similar datasets for select cities of China (30), including Guangzhou (31), Beijing (32), and European cities (as part of the project MapUCE, (33)). These datasets and the method for deriving UCPs using intensive manual intervention, development, and processing of high-resolution datasets. Typically, building footprints and individual building heights which are difficult to obtain.

With the increasing urbanization in developing countries, there is a growing need to generate UCPs more widely. Towards that objective, we introduce an automatic method for generating UCPs using an urban visualization approach. The work builds on the foundational work of (34–36) that have been aligned with the WUDAPT (37, 38) initiative. This approach utilizes deep learning combined with procedural modeling to infer various urban features despite only having limited information, and then automatically generates a 3D city model and its UCPs. The translation of such synthetic data, into a weather modeling framework would potentially open an avenue for developing simulations for locales where such measurements are lacking—which is more of a norm than the exception.

Accordingly, the study objectives described here are: (1) to introduce a novel automatic method to generate UPCs, (2) to demonstrate the development of a UCP dataset for Chicago, USA, and (3) to integrate these UCPs with the WRF model and evaluate the performance of the WRF model using these derived UCPs.

Methods

Digital synthetic city generation

We have developed a novel deep-learning and procedural modeling based method for creating a city-scale 3D urban model, called in this paper Digital Synthetic City (DSC), from which we can derive various urban morphology parameters. Our method uses satellite imagery and global-scale population and elevation data as input to our automatic method for producing a statistically
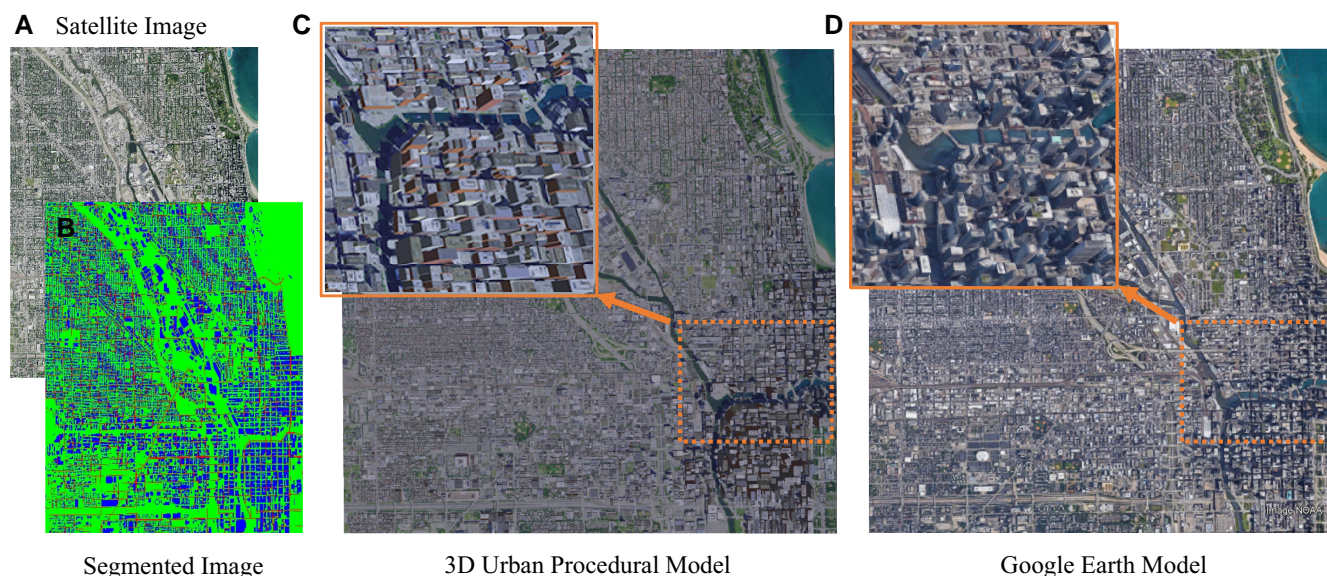
Fig. 1. A preliminary automatic synthetic modeling of Chicago, obtaining a detailed procedural model that is statistically similar to reality (Google Earth shown as reference).

similar and synthetic city-scale 3D urban model as output. The result is the ability to almost instantly create a plausible synthetic large-scale 3D urban model (Fig. 1).

The approach takes as input various geospatial products, summarized in Table 1. It consists of three main components: (1) building and parcel area estimation, (2) procedural model generation, and (3) an optional procedural model optimization (38). As shown in Fig. 2, we first utilize an image segmentation network (i.e. U-NET (42, 43)) and then a novel upsampling and sharpening network based on an autoencoder framework (44). Further, we combine building segmentation with a building setback prediction network. An optional optimization step uses information about the number and height of a few percent of buildings in the target area to calibrate the generated city to the target location. The end result is the ability to segment and infer building footprints accurately despite the relatively low-resolution of satellite imagery and the occlusions by nearby structures. The output of our method is a large spatial procedural city model consisting of 3D buildings distributed over the target area and registered in place with the road network, suitable for modeling urban areas worldwide for urban design in city planning and simulation. The method is leap-frog generational improvement from (34, 37) that develops a novel framework by considering the globally available datasets and upscaling method shown in Fig. 2. As a result, the similarity between the real and estimated morphological parameters is maintained. More information on DSC method is provided in the online supplementary material text and Figures S1–S4.

Estimation of urban canopy parameters

The subset of UCPs under our current consideration, which are the typical main parameters of urban areas modeled by systems such as WRF, are:

1. Building Height (Z_R):

$$Z_R = \frac{\sum_{i=1}^{N} A_i h_i}{\sum_{i=1}^{N} A_i} \tag{1}$$

Table 1. Summary of data sources used in the DSC generation.

Data name	Data source	Resolution	Scale
Road vector	Open street maps (39)	—	Most cities
Elevation data	JAXA (40)	30 m	Global
Population data	LandScan (41)	1 km	Global
Satellite data	PlanetScope	3 m	Global

where A_i is the plan area of the building, h_i is the height of the building, N is the total number of buildings within a given region.

2. Standard deviation of building height (σ_z):

$$\sigma_z = \sqrt{\frac{\sum_{i=1}^{N} (h_i - Z_R)^2}{N - 1}} \tag{2}$$

3. Roof width (W_{roof}): calculated by assuming buildings are rectangles of equal area (A) and perimeter (P) as of building footprint.

$$W_{roof} = \frac{P - \sqrt{P^2 - 16A}}{4} \tag{3}$$

4. Urban fraction (f_{urb}):

$$f_{urb} = \frac{\sum_{i=1}^{N} A_i}{A} \tag{4}$$

where A_i is the area of the ith building footprint for a given LCZ, A is the total area of the given LCZ.

5. Building height percentage bins: for each LCZ, buildings are placed into bins based on their height with a granularity of 5m.

Table 2 shows the difference between the WUDAPT and DSC-derived values of UCPs. The height percentages per bin amongst the various LCZ classes are shown in Figures S6 and S7. The spatial plots of urban fraction, building height, and road width is shown in Figure S8.

Table 2. WUDAPT and DSC-derived urban morphological parameters.

Parameters	f_{urb}	f_{urb}v	Z_R (m)	Z_R (m) (DSC)	W_{roof} (m)	W_{roof} (m) (DSC)
LCZ 1	1	0.81	37.5	30.21	15	19.14
LCZ 2	0.95	0.80	17.5	23.15	12.7	9.61
LCZ 3	0.9	0.77	6.5	19.38	5.7	5.40
LCZ 4	0.65	0.50	37.5	24.62	37.5	6.92
LCZ 5	0.7	0.79	17.5	25.67	33.3	5.89
LCZ 6	0.65	0.71	6.5	20.93	12.4	5.89
LCZ 7	0.85	—	3	—	2	—
LCZ 8	0.85	0.61	6.5	16.60	32.5	10.50
LCZ 9	0.3	—	6.5	-	10	—
LCZ 10	0.55	0.70	10	27.97	28.5	21.84

Currently, LCZ 7 (lightweight low-rise) and 9 (sparsely built) do not have DSC-derived values, hence WUDAPT values are used as default here.
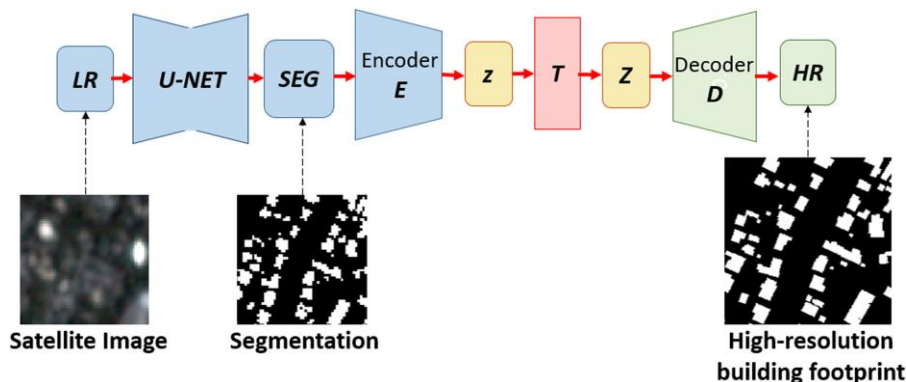

Fig. 2. Segmentation and upsampling to obtain detailed building footprints, which are later converted to 3D building envelope models.
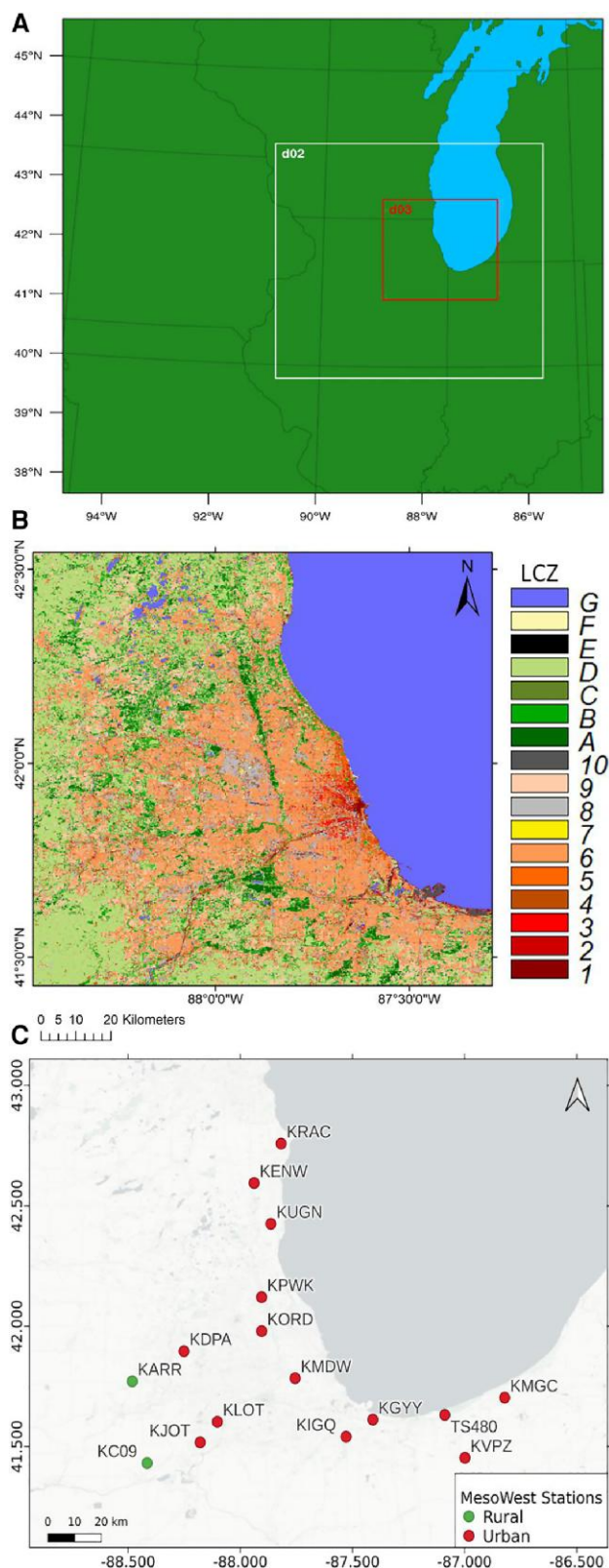
Fig. 3. (A) WRF domain configuration. (B) Local Climate Zone map for Chicago, USA. (C) Location of Mesowest observation over rural and urban regions.

Modeling experiments and evaluation

One of the challenges that exist when creating such a high-resolution data set is, how to verify the output? It is important to highlight that DSC is not an exact replication of the urban

morphology. In fact, the efficiency of the DSC framework lies in the flexibility to create variable grid resolution (spacing) urban morphological parameters in a "fast" manner (matter of minutes). The DSC output need to be evaluated for "fit for purpose" and not just the geometric reproducibility of a corresponding Google Earth or similar available dataset in the public domain.

Therefore, for assessing the suitability of DSC for urban modeling studies, we design a modeling experiment for a real-world case focusing on the Chicago downtown region using the WRF model.

DSC-WRF urban modeling

The simulations were performed using the Weather Research and Forecasting (WRF) model, version 4.2.1 (45). A typical model configuration is considered. Fig. 3A shows the three nested domains centered over Chicago, USA, with a spatial resolution of 9, 3, and 1 km for outermost, middle and inner most domains, respectively. The model was configured with 42 pressure levels with first model level located at 21.2 m and first 1 km vertical height containing 11 model levels. The initial and boundary conditions are taken from the National Centers for Environmental Prediction (NCEP) Final Reanalysis dataset at 1×1 degree spatial and 6-hourly temporal resolution. The physics components include: the WRF single moment 6 class (46) for microphysics, (47) for shortwave, the Rapid Radiative Transfer Model for longwave radiation parameterizations (48), (49) for the planetary boundary layer, Noah (50) for land surface model, Building Environment Parametrization (BEP) (8) for the urban model, and (51) for the cumulus scheme (only for the outermost domain of 9 km spatial resolution). The LCZs for Chicago, USA, are generated using the crowd-sourcing method of (52). The training dataset, created manually, is obtained from the WUDAPT portal and random forest classification is applied to Landsat 8 imagery to derive the LCZs for the desired region. The final map is shown in Fig. 3B. The simulations are performed from 1/Jul/2018 00:00 to 7/Jul/2018 06:00 UTC where first 6 h are discarded as spin-up time.

The modeling experiments are divided into three simulations based on the source of land use/land cover and UCP values: (1) NUDAPT (Control), (2) WUDAPT, (3) DSC. The Control simulations uses National Land Cover Database land use/land cover with NUDAPT parameters, the three default WRF urban classes and corresponding UCPs. The WUDAPT uses the MODIS classes with additional urban LCZs and UCPs from (53); while. the DSC uses the WUDAPT classes with UCPs generated from our automatic method. The WRF simulations are evaluated using the 16 Mesowest stations located in the innermost domain (shown in Fig. 3C). The root mean square error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient (r) against the observational 2 m air temperature, 2 m relative humidity, and 10 m wind speed data from these stations are used to evaluate the different simulations. The details of Mesowest observation stations are provided in Table S1 and overview of methodology is shown in Figure S5. The urban heat island intensity is calculated by subtracting urban to rural 2 m air temperatures.

The 2018 July 1–7, period represent the weather over Chicago, USA, after the hottest day (June 30) since 2012. On July 1, the temperatures start to face until July 3, as the cold front departs through the region and a surface ridge shifts towards the east of the region. From July 3–5, the temperatures again start rising due to the moist and warm air mass over the area. On July 5, the heat and humidity help support the initiation of isolated and scattered thunderstorms in the region. Finally, on July 6–7, the thunderstorms, the movement of the cold front, and the advection

Table 3. WRF Model evaluation for Control, WUDAPT, and DSC simulations.

	Variable	Simulation	RMSE	MAE	*r*
Day (05:00–20:00 LT)	2 m air temperature (K)	Control	3.34	2.61	**0.67**
		WUDAPT	3.25	2.45	0.65
		DSC	**3.18**	**2.42**	0.66
	2 m specific humidity (g/kg)	Control	**2.38**	**1.76**	**0.80**
		WUDAPT	2.62	1.91	0.77
		DSC	2.62	1.89	0.77
	10 m wind speed (m/s)	Control	**1.80**	**1.40**	**0.57**
		WUDAPT	1.83	1.42	0.57
		DSC	1.98	1.56	0.47
Night (20:00–05:00 LT)	2 m air temperature (K)	Control	3.32	2.30	0.46
		WUDAPT	3.12	2.28	0.54
		DSC	**3.07**	**2.25**	**0.54**
	2 m specific humidity (g/kg)	Control	2.57	1.83	0.68
		WUDAPT	2.26	1.65	0.76
		DSC	**2.20**	**1.60**	**0.77**
	10 m wind speed (m/s)	Control	1.84	1.46	0.29
		WUDAPT	1.72	1.36	0.37
		DSC	**1.54**	**1.20**	**0.38**
All (00:00–23:00 LT)	2 m air temperature (K)	Control	2.43	1.85	**0.86**
		WUDAPT	**2.34**	**1.78**	0.85
		DSC	2.38	1.81	0.84
	2 m specific humidity (g/kg)	Control	**1.81**	**1.39**	0.87
		WUDAPT	1.85	1.44	0.87
		DSC	1.85	1.41	0.87
	10 m wind speed (m/s)	Control	1.67	1.29	0.61
		WUDAPT	**1.63**	**1.26**	**0.63**
		DSC	1.71	1.32	0.56

Bold text represents best score. *r* is statistically significant (P-value <0.05).

from lake breeze towards the urban area allowed the temperatures to reduce.

Results and discussions

The notable weather features for 2018 July 1–7, over Chicago, USA, shows the temperature variations from 290 to 305 K, which contain typical urban heat island feedbacks and land-lake breeze circulation. Therefore, we discuss the results of the simulations for three key variables, temperature, relative humidity, and wind (speed and direction), that are impacted by urban structures.

Evaluation of WRF simulations

The performance of the WRF simulations is evaluated for day (sunrise to sunset; 05:00–20:00 LT), night (20:00–05:00 LT), and whole period (All) consisting of day and night (see Table 3). During the daytime, DSC outperforms WUDAPT and Control simulations for 2 m air temperature. However, 2 m specific humidity and 10 m wind speed is better simulated in the Control. The reduced 10 m wind speed in case of DSC is due to increased roughness length (taller buildings) relative to WUDAPT and Control. The correlation coefficient during the daytime is relatively better for Control simulation. During the nighttime, 2 m air temperature, 2 m specific humidity, and 10 m wind speed is better simulated by the DSC simulations. Comparing the entire period, WUDAPT shows performance better for 2 m air temperature and 10 m wind speed, while 2 m specific humidity is better simulated by Control. Thus, the mean statistics suggest that WUDAPT and DSC perform better than Control, and a negligible difference is noted between WUDAPT and DSC simulations. The DSC values are closer to the WUDAPT simulations for the entire simulation period (for latent, sensible heat fluxes, and friction velocity see Fig. S10). This highlights that the UCP values generated from the DSC are performing similar to the literature derived values,

bolstering confidence in the automated supporting the methodology adopted in this study.

Here, it is important to highlight that the evaluation represent the effect of urban morphology by providing bulk values to LCZs. High-resolution gridded datasets such as (54, 55) are expected to provide better urban morphology. There are other sources of uncertainty such as the lack of representation of street trees and urban vegetation (56). Other UCPs (such as albedo, emissivity, and thermal properties of the building material) that affects the local climate zones are not considered, thus adding uncertainty in the simulations. However, the automated city-scale urban morphology generator framework adopted for urban weather modeling is novel and effective for regional studies.

Diurnal and urban heat island intensity

The time-series of the variables (shown in Fig. 4A,C,E) shows that all simulations follow the observations until 2018 July 3, where the 2 m air temperature and 10 m wind speed drops to less than 290 K and 1 ms^{-1} respectively. On 2018 July 5, the increase in the 2 m specific humidity, reduction in 2 m air temperature is not well captured in the simulations. The simulation diurnal profile of the variables is shown in Fig. 4B,D,F. The Control simulation overestimates the afternoon 2 m air temperature by ≈1.5 K. The DSC and WUDAPT simulations show consistent behavior for 2 m air temperature and specific humidity and are closer to the observations. During the daytime, DSC shows a reduced 10 m wind speed while WUDAPT and Control simulations are closer to the observations. For, nighttime 10 m wind speed is better captured by the DSC simulations. The difference between the rural and urban temperature is shown in the form of urban heat island intensity (UHII, rural and urban stations shown in Fig. 3C) (see Fig. 5). The rural area starts warming from sunrise (05:00 LT) to afternoon (14:00 LT), while a counter effect is observed in the day to night transition, rural is cooling faster than the urban areas. Thus,
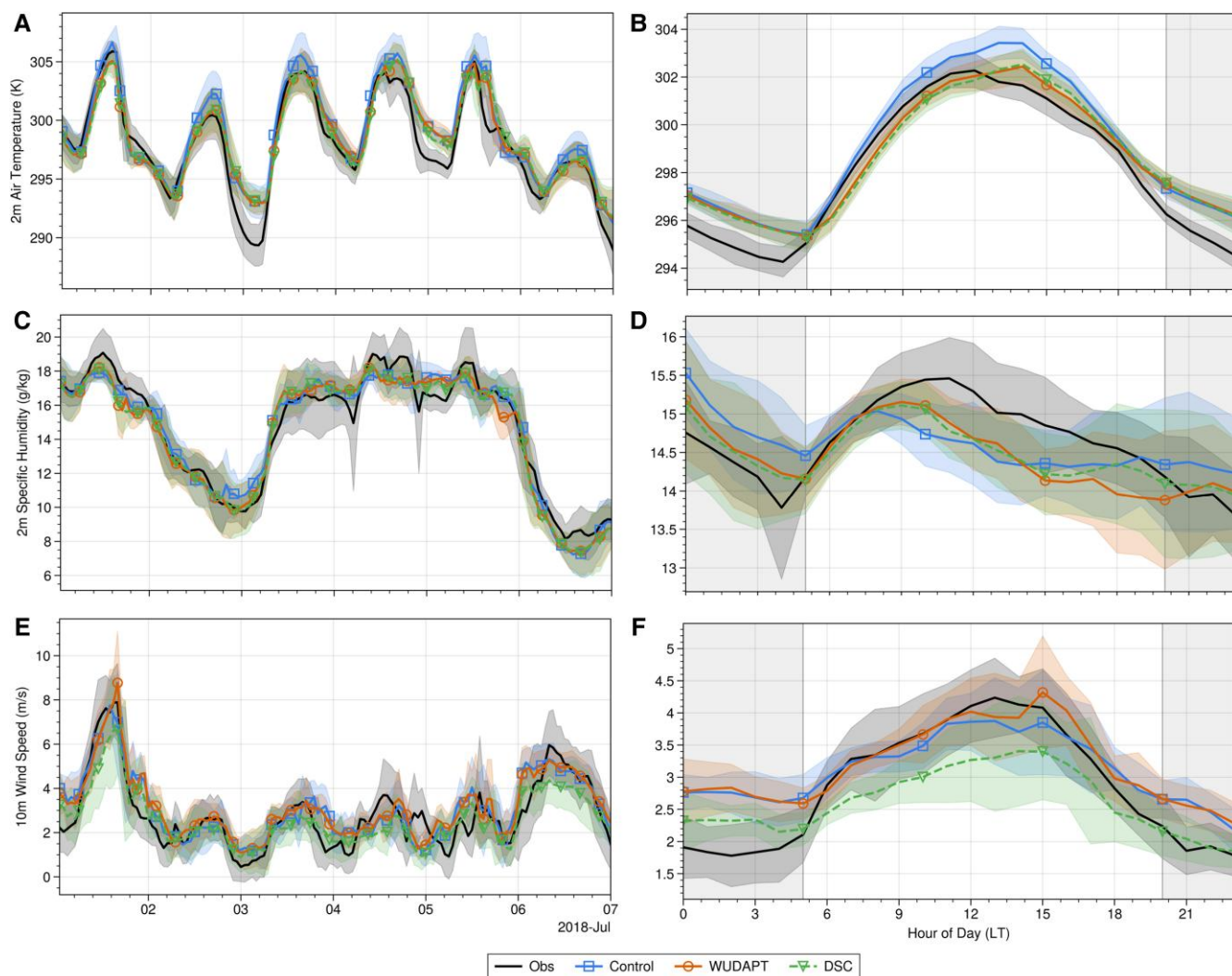
Fig. 4. Mean of all stations for (A) 2 m air temperature, (C) 2 m specific humidity, and (E) 10 m wind speed from observations (Obs, solid line), Control (square marker solid line), WUDAPT (circle marker solid line), and DSC (triangle dashed line) simulations. Similarly, the composite diurnal cycle is shown in (B) 2 m air temperature, (D) 2 m specific humidity, and (F) 10 m wind speed. The shaded portions are ±1 standard deviation.
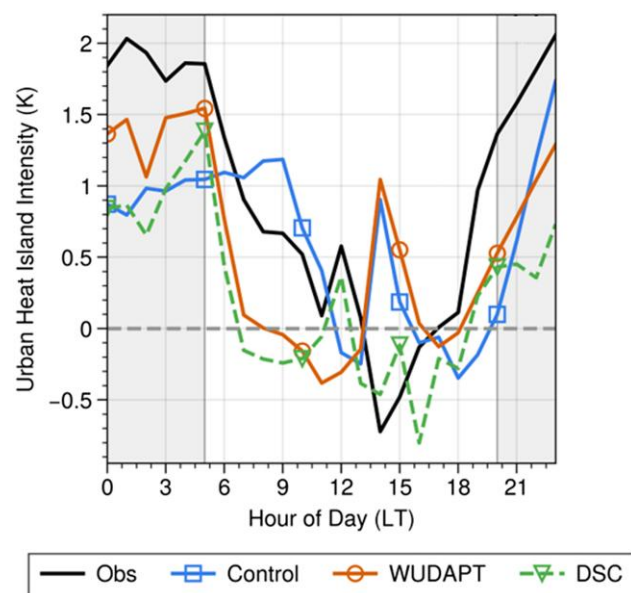

Fig. 5. Urban heat island intensity from Obs, Control, WUDAPT, and DSC simulations.

higher (>2 K) urban heat island intensity is noted for nighttime. DSC simulations captures afternoon urban cooling but underestimates the nighttime UHII by ≈1 K. The reduction in the UHI may be attributed to the reduced urban fraction in DSC as compared to the WUDAPT. Overall, changing the UCPs from WUDAPT to DSC has non-linear feedback leading to changes in simulated weather inside and outside the city. Changes in the temperature of each LCZ and spatially are shown in Figures S11 and S12, respectively. The kernel density estimates of 2 m specific humidity, 2 m air temperature, and 10 m wind speed are also provided in Figure S9.

Wind speed and direction

Figure 6 shows the wind direction and speed for the Control simulations. The wind speed within the city is lower than the surrounding areas, during all times of the day due to relatively high roughness lengths of the city. The WUDAPT simulation shows a negligible change in the wind speed within the city when compared with the Control run (Fig. 6B,F,J,N). The DSC simulation shows a reduced wind speed from all the simulations. This change in wind speed has a small but notable non-linear effect on the
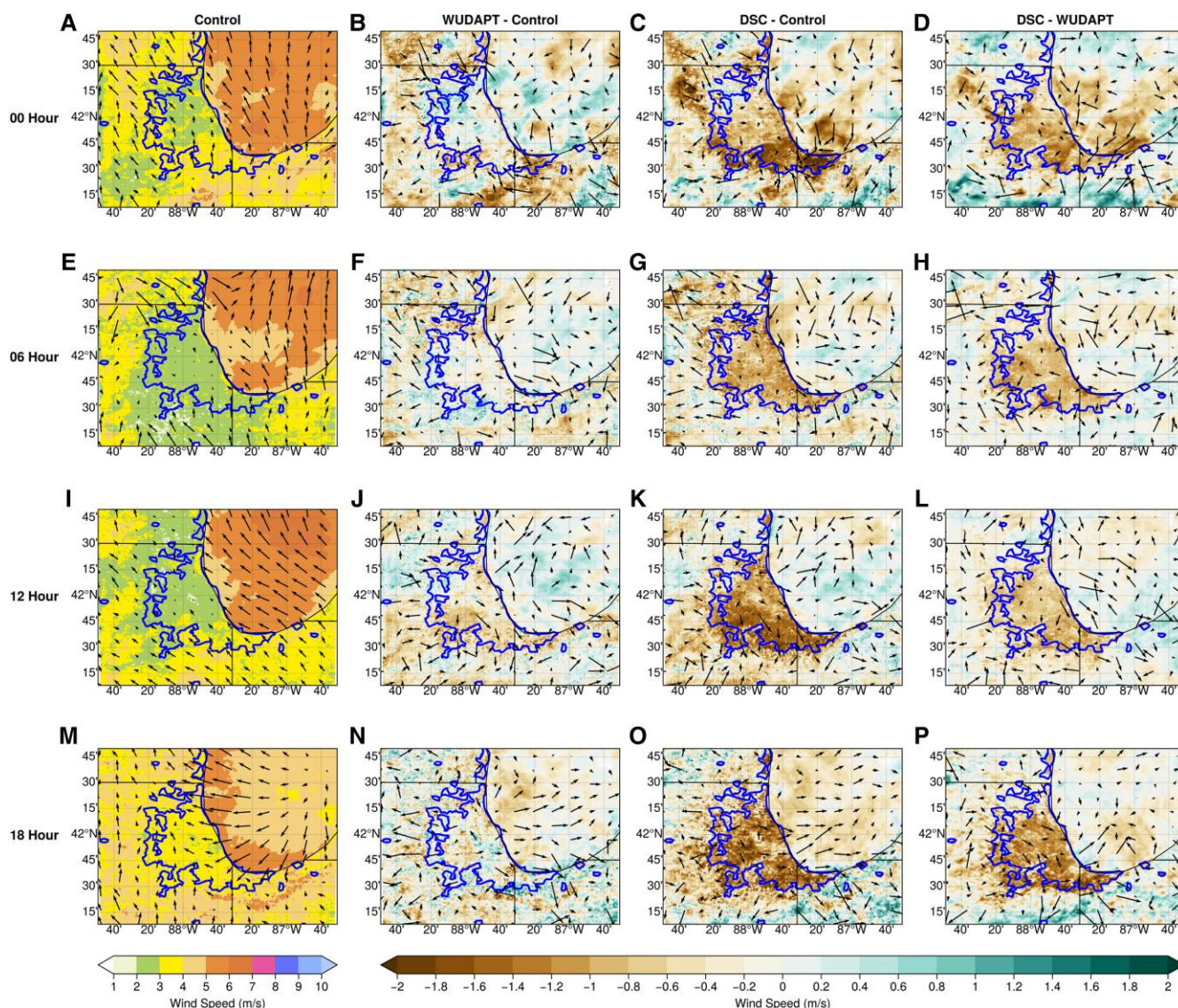
Fig. 6. 10 m wind speed (background) (ms^{-1}) and direction at 00- (A–D), 06- (E–H), 12- (I–L), and 18- (M–P) hour average for the Control simulation (A,E,I,M). Difference between the 10 m wind speed (background) and direction of WUDAPT and Control (positive indicates WUDAPT is larger) (B,F,J,N), DSC and Control (C,G,K,O), and DSC and WUDAPT (D,H,L,P).

atmospheric circulation dynamics, such as the land-sea breeze circulation. The relatively lesser change of 2 m temperature than the 10 m wind speed can be attributed to the significant difference in the building heights and roof widths that affects the roughness lengths, thus modulating the model outputs (see Fig. S8; (8)). Similar results were also observed by Wang et al. (57) and Loridan et al. (21) for offline simulations of the single-layer UCM. Additionally, the results from the Control simulations are in parity with (58).

Conclusions

UCPs are an important component of urban climate modeling. This study introduces a new methodology to calculate city-wide UCPs using an automatic deep learning-based synthetic data generation framework from globally available products. The newly generated UCPs utilized for environmental simulations in Chicago, USA using the WRF model. A total of three simulations consisting of a Control run (using NUDAPT dataset), WUDAPT

(incorporation of LCZs), and DSC (using LCZs and new UCPs) were conducted. The results show that urban LCZs have a significant impact on the simulation of air temperature. Moreover, the automatically computed DSC parameter values yield simulation results as good as, or sometimes more accurate, than WUDAPT (which requires crowd-sourcing and benefits from a hand-crafted dataset optimization). The changes in the UCPs also impacted the overall simulations by reducing the wind speed (due to increased roughness length) within the urban area and small changes in the temperature values (due to urban fraction). Thus, the automation rendered by the DSC method opens the opportunity to a more scalable and widespread ability to perform more accurate urban meteorological modeling and forecasting.

An overarching conclusion, the DSC renders a visualization of the urban canopy by producing urban structures/environment details that can be used in representing the urban areas within the UCMs. As future work, we see three avenues. Firstly, we would like to extend our DSC method to include support for all LCZ classes, potentially leading to increased accuracy. Secondly, we

would like to improve the accuracy of determining parameters for mostly green areas within the city and in the para-urban region. This may have a significant effect on temperature and humidity estimates. Lastly, we would like to tie our synthetic generation ability with urban planning policies so that what-if scenarios can be generated based on desired urban meteorological consequences.

Acknowledgments

Supplementary material

Supplementary material is available at *PNAS Nexus* online.

Funding

Authors' contributions

D.A. and D.N. conceived the experiment(s), P.P., R.K., and L.H. conducted the experiment(s), P.P., R.K., and L.H. analysed the results. P.P., R.K., L.H., D.A., and D.N. wrote and reviewed the manuscript.

Data availability

The data underlying this article are available in Zenodo at 10.5281/zenodo.7077486.

References

1   UN. 2021. Cities – united nations sustainable development.

2   Wahl T, Jain S, Bender J, Meyers SD, Luther ME. 2015. Increasing risk of compound flooding from storm surge and rainfall for major us cities. *Nat Clim Change*. 5(12):1093–1097.

3   Guerreiro SB, Dawson RJ, Kilsby C, Lewis E, Ford A. 2018. Future heat-waves, droughts and floods in 571 European cities. *Environ Res Lett*. 13(3):034009.

4   Liu J, Niyogi D. 2019. Meta-analysis of urbanization impact on rainfall modification. *Sci Rep*. 9(1):1–14.

5   Rizwan AM, Dennis LYC, Liu C. 2008. A review on the generation, determination and mitigation of urban heat island. *J Environ Sci*. 20(1):120–128.

6   Mage D, *et al.* 1996. Urban air pollution in megacities of the world. *Atmos Environ*. 30(5):681–686.

7   Chen F, *et al.* 2011. The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *Int J Climatol*. 31(2):273–288.

8   Martilli A, Clappier A, Rotach MW. 2002. An urban surface exchange parameterisation for mesoscale models. *Boundary Layer Meteorol*. 104(2):261–304.

9   Salamanca F, Martilli A. 2010. A new building energy model coupled with an urban canopy parameterization for urban climate simulations—Part II. Validation with one dimension offline simulations. *Theor Appl Climatol*. 99(3):345–356.

10 Masson V, *et al.* 2013. The surfexv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geosci Model Dev*. 6(4):929–960.

11 Oleson W, Bonan B, Feddema J, Vertenstein M, Kluzek E. 2010. Technical description of an urban parameterization for the community land model (CLMU). Boulder: NCAR.

12 Patel P, Karmakar S, Ghosh S, Niyogi D. 2020. Improved simulation of very heavy rainfall events by incorporating WUDAPT urban land use/land cover in WRF. *Urban Clim*. 32:100616.

13 Kusaka H, Kondo H, Kikegawa Y, Kimura F. 2001. A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Boundary Layer Meteorol*. 101(3):329–358.

14 Oke TR, Mills G, Christen A, Voogt JA. 2017. *Urban climates*. Cambridge: Cambridge University Press.

15 Järvi L, Grimmond CSB, Christen A. 2011. The surface urban energy and water balance scheme (SUEWS): evaluation in Los Angeles and Vancouver. *J Hydrol (Amst)*. 411(3-4):219–237.

16 Harshan S, Roth M, Velasco E, Demuzere M. 2018. Evaluation of an urban land surface scheme over a tropical suburban neighborhood. *Theor Appl Climatol*. 133(3):867–886.

17 Schmid PE, Niyogi D. 2017. Modeling urban precipitation modification by spatially heterogeneous aerosols. *J Appl Meteorol Climatol*. 56(8):2141–2153.

18 Hosannah N, Gonzalez JE. 2014. Impacts of aerosol particle size distribution and land cover land use on precipitation in a coastal urban environment using a cloud-resolving mesoscale model. *Adv Meteorol*. 2014:904571.

19 Wouters H, *et al.* 2016. Efficient urban canopy parametrization for atmospheric modelling: description and application with the COSMO-CLM model (version 5.0 _clm6) for a Belgian summer. *Geosci Model Dev*. 9:3027–3054.

20 Powers JG, *et al.* 2017. The weather research and forecasting model: overview, system efforts, and future directions. *Bull Am Meteorol Soc*. 98(8):1717–1737.

21 Loridan T, *et al.* 2010. Trade-offs and responsiveness of the single-layer urban canopy parametrization in WRF: an offline evaluation using the MOSCEM optimization algorithm and field observations. *Q J R Meteorol Soc*. 136(649):997–1019.

22 Salamanca F, Martilli A, Tewari M, Chen F. 2011. A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF. *J Appl Meteorol Climatol*. 50(5):1107–1128.

23 Li D, Bou-Zeid E, Baeck ML, Jessup S, Smith JA. 2013. Modeling land surface processes and heavy rainfall in urban environments: sensitivity to urban surface representations. *J Hydrometeorol*. 14(4):1098–1118.

24 Carter M, Shepherd JM, Burian S, Jeyachandran I. 2012. Integration of lidar data into a coupled mesoscale–land surface model: a theoretical assessment of sensitivity of urban–coastal mesoscale circulations to urban canopy parameters. *J Atmos Ocean Technol*. 29(3):328–346.

25 Ching J, *et al.* 2018. WUDAPT: an urban weather, climate, and environmental modeling infrastructure for the anthropocene. *Bull Am Meteorol Soc*. 99(9):1907–1924.

26 Stewart ID, Oke TR. 2012. Local climate zones for urban temperature studies. *Bull Am Meteorol Soc.* 93(12):1879–1900.

27 Stewart ID, Oke TR, Krayenhoff ES. 2014. Evaluation of the "local climate zone" scheme using temperature observations and model simulations. *Int J Climatol.* 34(4):1062–1080.

28 Brousse O, *et al.* 2019. Using local climate zones in sub-saharan Africa to tackle urban health issues. *Urban Clime.* 27:227–242.

29 Ching J, *et al.* 2009. National urban database and access portal tool. *Bull Am Meteorol Soc.* 90(8):1157–1168.

30 Li H, *et al.* 2022. Improving the WRF/urban modeling system in China by developing a national urban dataset. *Geosci Front.* 13(4):101385.

31 Shen C, *et al.* 2019. Impacts of high-resolution urban canopy parameters within the WRF model on dynamical and thermal fields over Guangzhou, China. *J Appl Meteorol Climatol.* 58(5):1155–1176.

32 He X, *et al.* 2019. High-resolution dataset of urban canopy parameters for Beijing and its application to the integrated WRF/urban modelling system. *J Clean Prod.* 208:373–383.

33 Masson V. 2015. Urban climate, human behavior & energy consumption: from LCZ mapping to simulation and urban planning (the MApUCE project). In: 9th International Conference on Urban Climate, Toulouse, France.

34 Aliaga DG, Vanegas C, Lei M, Niyogi D. 2013. Visualization-based decision tool for urban meteorological modeling. *Environ Plann B: Plann Des.* 40(2):271–288.

35 Yang L, *et al.* 2016. Contrasting impacts of urban forms on the future thermal environment: example of Beijing metropolitan area. *Environ Res Lett.* 11(3):034018.

36 Garcia-Dorado I, Aliaga DG, Bhalachandran S, Schmid P, Niyogi D. 2017. Fast weather simulation for inverse procedural design of 3D urban models. *ACM Trans Graph (TOG).* 36(2):1–19.

37 Ching J, *et al.* 2019. Pathway using WUDAPT's digital synthetic city tool towards generating urban canopy parameters for multi-scale urban atmospheric modeling. *Urban Clim.* 28:100459.

38 Zhang X, Shehata A, Benes B, Aliaga D. 2020. Automatic deep inference of procedural cities from global-scale spatial data. *ACM Trans Spat Algorithms Syst.* 7(2):1–28.

39 Haklay M, Weber P. 2008. Openstreetmap: user-generated street maps. *IEEE Pervasive Comput.* 7(4):12–18.

40 Nikolakopoulos KG. 2017. Evaluating ALOS AW3D30 data. In: Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017). Vol. 10444. International Society for Optics and Photonics. p. 1044408.

41 Bright EA, Coleman PR, Rose AN, Urban ML. 2012. Landscan 2011. Oak Ridge, TN: Oak Ridge National Laboratory SE.

42 Ronneberger O, Fischer P, Brox T. 2015. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer. p. 234–241.

43 Iglovikov V, Seferbekov S, Buslaev A, Shvets A. 2018. Ternausnetv2: fully convolutional network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA. p. 233–237.

44 Kingma DP, Welling M. 2013. Auto-encoding variational Bayes, arXiv, arXiv:1312.6114, preprint: not peer reviewed.

45 Skamarock WC, *et al.* 2019. A description of the advanced research WRF model version 4. Boulder, CO, USA: National Center for Atmospheric Research. p. 145.

46 Hong S-Y, Lim J-OJ. 2006. The WRF single-moment 6-class microphysics scheme (WSM6). *Asia Pac J Atmos Sci.* 42(2):129–151.

47 Dudhia J. 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J Atmos Sci.* 46(20):3077–3107.

48 Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J Geophys Res Atmos.* 102(D14):16663–16682.

49 Bougeault P, Lacarrere P. 1989. Parameterization of orography-induced turbulence in a mesobeta–scale model. *Mon Weather Rev.* 117(8):1872–1890.

50 Tewari M, *et al.* 2004. Implementation and verification of the unified NOAH land surface model in the WRF model. In: 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, Washington, USA. Vol. 1115.

51 Grell GA, Dévényi D. 2002. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys Res Lett.* 29(14):38–1.

52 Bechtel B, *et al.* 2015. Mapping local climate zones for a world-wide database of the form and function of cities. *ISPRS Int J Geoinf.* 4(1):199–219.

53 Brousse O, Martilli A, Foley M, Mills G, Bechtel B. 2016. WUDAPT, an efficient land use producing data tool for mesoscale models? Integration of urban LCZ in WRF over Madrid. *Urban Clim.* 17:116–134.

54 Zhang X, *et al.* 2020. Modelling urban meteorology with increasing refinements for the complex morphology of a typical Chinese city (Xi'an). *Build Environ.* 182:107109.

55 Li H, Liu Y, Zhang H, Xue B, Li W. 2021. Urban morphology in China: dataset development and spatial pattern characterization. *Sustain Cities Soc.* 71:102981.

56 Mussetti G, *et al.* 2020. COSMO-BEP-Tree v1. 0: a coupled urban climate model with explicit representation of street trees. *Geosci Model Dev.* 13(3):1685–1710.

57 Wang Z-H, Bou-Zeid E, Au SK, Smith JA. 2011. Analyzing the sensitivity of WRF's single-layer urban canopy model to parameter uncertainty using advanced monte carlo simulation. *J Appl Meteorol Climatol.* 50(9):1795–1814.

58 Sharma A, *et al.* 2017. Urban meteorological modeling using WRF: a sensitivity study. *Int J Climatol.* 37(4):1885–1900.