

Computers, Environment and Urban Systems
Simulating large-scale urban land-use patterns and dynamics using the U-Net deep learning architecture
--Manuscript Draft--

Manuscript Number:	CEUS-D-21-00877R1
Article Type:	Research Paper
Keywords:	Deep Learning; urban expansion simulation; cellular automata; spatial pattern
Corresponding Author:	jinzhu wang Deakin University - Melbourne Burwood Campus AUSTRALIA
First Author:	jinzhu wang
Order of Authors:	jinzhu wang Michalis Hadjikakou Richard J. Hewitt Brett A. Bryan
Abstract:	Accurate predictions of land-use change are important for supporting planning. Cellular automata (CA) models are widely used to simulate real-world urban land-use change but accurately modelling complex spatial urban patterns and dynamics can be challenging due to the high degree of subjectivity involved in CA model parameterisation. Advances in deep learning enable complex spatial patterns such as urban development to be learned and simulated. In this study, we used the U-Net deep learning algorithm to capture historical urban development and simulate future patterns for the North China Plain, one of the most rapidly urbanizing regions on the planet. We validated the model against a reference map for 2018 then applied it in predicting patterns of urban expansion for 2030. The results showed that U-Net can accurately predict urban land-use and mimic real-world spatial patterns with very low requirement for model parameterization and forcing data. Visual inspection of the outputs revealed that U-Net was able to automatically learn complex urban development patterns and processes such as neighbourhood influence, the gravity effects of large cities, and the tendency for linear development. Deep learning architectures such as U-Net provide a new parameter-free way to accurately capture and simulate spatial features in projections of future urban development and land-use change.
Suggested Reviewers:	Haijun Wang landgiswhj@163.com He integrated historical urban development with CA models to simulate urbanization, which enables him to judge my paper from the perspective of innovative urban simulation. Keith C. Clarke kcclarke@ucsb.edu He calibrated the SLEUTH model to project land use to 2100. So he is capable of criticising my research regarding the improvement of urban simulation. Markus Reichstein mreichstein@bgc-jena.mpg.de He published a paper on geospatial big data and machine learning on Nature, he can evaluate my study of using deep learning to simulate urban land use.
Opposed Reviewers:	
Response to Reviewers:	



26th May 2022

To the Editor, Computers, Environment and Urban Systems

Dear Professor Alison Heppenstall,

Re: Revision of CEUS-D-21-00877 Wang et al. "Simulating large-scale urban land-use patterns and dynamics using the U-Net deep learning architecture."

We are very grateful for your interest in our work. We greatly appreciate the valuable comments and suggestions from the reviewers.

We are delighted that reviewers broadly appreciated the significance of our work on introducing deep learning to the simulation of urban land-use change. We note that Reviewer #1 said that "the subject addressed in this article is worthy of investigation," and Reviewer #2 thought "it is a promising new idea to apply U-Net to urban growth simulation and prediction". We also appreciate the Reviewers' more critical comments and suggestions, which helped us to greatly improve the novelty, reliability, and impact of our research.

Reviewer #1 suggested that we clarify the data used in the study and the description of the two trained models. This reviewer also suggested a consistent use of "urban land-use" throughout the manuscript and several specific modifications regarding the figure layouts, the use of adjectives, etc. In response, we added a table to describe the data used in this study, changed the various phrases *urban dynamic map*, *urban map*, *urban pixels*, and *urban images* to *urban land-use*, and added more descriptions to the figures.

Reviewer #2 advised us to discuss how deep learning helps the understanding of urban land-use change mechanisms and if deep learning had more robust performance than CA-based methods. In response, we added a discussion on techniques to transform deep learning structures into human-recognizable knowledge. Meanwhile, we emphasized that the primary goal of introducing deep learning to urban land-use simulation is to explore how deep learning could complement, rather than replace, CA-based land-use modeling. We also modified the discussion on comparing deep learning with CA to clarify that the U-Net had learned urban land-use patterns and achieved accuracies similar to CA-based models conducted in a similar urban growth context.

We have addressed each comment one by one as detailed in the attached "Response to Reviewers" document. All comments are reproduced and our responses are given directly afterward in a different color (blue). We have made our best effort to address each comment. We believe that these modifications will significantly increase the novelty, reliability, and impact of our research. We kindly ask that if, after considering our responses, the Reviewers

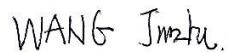
and Editors still have further comments on our manuscript, we remain very open to further discussion and revision of the manuscript.

With warm regards.

Yours Sincerely,

Jinzhu WANG, on behalf of all authors

Sincerely, and with warm regards



Ph.D. candidate,

Planet-A Sustainability Science

School of Life and Environmental Sciences

Deakin University

Melbourne Campus, 221 Burwood Hwy, Burwood, VIC 3125

Ph: +61 (0) 481 280 877 | [Google Scholar](#) | [YouTube](#) |

Reviewer #1:

The paper reports the result of a study using deep learning to model urban land development by applying the U-Net architecture to a study area of the North China Plain. The subject addressed in this article is worthy of investigation and appropriate to the journal.

Thanks for confirming the value of this study.

However, the methodology is not clearly described w.r.t the data used for the development of the two U-Net models and the novel methodological contributions.

To better explain the data used to train the U-Net models, we added a table in the “Data processing” section.

Table 3. The data used to train U-Net models. (This table is inserted in the “2.4. Data preprocessing”)

Data	Time	Source	Resolution
Urban land use	1994, 2006, 2018	Wang et al. (2021b)	30 m
Elevation	2000*	Shuttle Radar Topography Mission	30 m
Slope	2000*	Derived from elevation	30 m

* Note, the elevation and slope data were used in 1994, 2006, and 2018 despite their being acquired in 2000.

4.1. Mimicking real-world urban patterns and dynamics

U-Net was able to model urban land-use change at high accuracy as well as capture and assimilate high-level spatial patterns of urban development in the North China Plain. First, the model captured neighborhood effects. The transition potential map revealed that lands near existing urban areas were more likely to transition to urban land-use in the future and that urbanization amongst larger expanses of non-urban land would be unlikely (Fig. 9). Second, U-Net was able to assimilate small-scale neighborhood effects into large-scale gravity effects. Larger areas of land near larger cities were predicted to become urbanized in the future, while much smaller areas were predicted to become urbanized around smaller, more remote villages (Fig. 10). Third, U-Net, rather than just simply buffering spatial features, captured the tendency of urban expansion to follow linear patterns. For example, the elongated development trend of Fengning County controlled by valley terrain was identified in the prediction (Fig. 12), and areas along major transport routes were allocated high transition potentials (Fig. 9). However, the U-Net model was unable to predict new planned developments that sprouted up some distance from existing urban areas, such as Binhai Zhen, Yangkou Zhen, Huying Zhen, and Xuancheng City (Fig. 10).

4.2. Reducing subjectivity by automatically constructing transition rules
The large-scale spatial patterns identified by U-Net, in particular the gravity effect of large cities, would not have been captured by existing CA models. A large neighborhood is required for a CA model to incorporate information at a large scale which tends to decrease simulation performance. Wang et al. (2021a) found that the best neighborhood size to simulate urban development in Beijing was 25 × 25 pixels, while Roodposhti et al. (2020) observed that a 9 × 9 pixel neighborhood outperformed other settings. However, the unique design of the U-Net model is capable of incorporating more extensive

information than previous CA model structures, enabling the neighborhood influence to be learned from the data and urban development to be simulated with more refined spatial configurations.

U-Net did not require extensive amounts of forcing data or parameterization to simulate urban land-use changes, except for the learning rate and the size of the training images. The transition rules were automatically generated in the form of learned weights, and training was an automatic process that required few subjective decisions to be made. In contrast, CA model structures have often required a complex calibration process to determine suitable parameters. For example, Feng and Tong (2020) developed a framework that integrated three algorithms and different neighborhood settings to simulate urban growth in Shanghai, whereas an array of parameters including inertia weights, decay magnitude, spatial heterogeneity, and variable scaling, needed to be specified before running the model. Many studies determine these parameters according to expert knowledge (Chen et al., 2020; Mustafa et al., 2017; Tripathy and Kumar, 2019), potentially leading to subjective biases. Other studies went through a systematic parameter selection process to find the best parameters (Roodposhti et al., 2020; Yu et al., 2021), but this process is time-consuming and impractical given the many unknown parameters.

Somehow, the conclusions are not well supported by the study results, e.g., "The U-Net model successfully captured neighborhood effects, gravity effects, and linear expansion along transportation routes in the urban dynamics", where "linear expansion along transportation routes" seems to be concluded based on visual inspections of 11 smaller areas.

We computed the patch number and landscape shape index to quantitatively assess the performance of the U-Net simulation. However, "neighborhood effects, gravity effects, and linear expansion effects" could not be quantified by current landscape shape indexes which have limited power to capture these high-level characteristics (Frazier and Kedron, 2017; Pontius et al., 2008; Uuemaa et al., 2009). Indeed, assessing the quality of output features remains a significant challenge even in the deep learning community because of the heterogeneity in spatial features that are beyond the ability of simple shape indexes to describe mathematically (Ghosh et al., 2020; Wang et al., 2020). Instead, it is common to evaluate the performances of deep learning models by observing the resulting images (Gonog and Zhou, 2019; Xu et al., 2018) and qualitatively assessing and categorizing the results via human interpretation to complement the quantitative metrics as we have done.

Our conclusions are not based on the 11 inset areas presented but rather they are based on detailed visual interpretation and query of the entire map outputs at a range of scales as we describe in the Methods section. The 11 small-scale (1km – 10km) regions were simply chosen to illustrate the urban patterns learned by U-Net (i.e., the neighborhood, gravity, and linear effects) to the reader. The illustrative inset areas were selected to cover the full gradient of city size and span the entire study area. The reason for choosing small-scale regions is that urban expansion simulation usually performs less accurately in small areas than in larger regions (Pontius et al., 2008). If the simulation in a small area looks sensible, readers are more likely to be convinced that the deep learning architecture could simulate the urban development well at the broader scale. The full dataset is available online

(<https://doi.org/10.6084/m9.figshare.19880671.v1>) if the reader wishes to make a more detailed inspection or inspect specific areas.

To ensure that readers are clear around the role and purpose of our visual inspection-based pattern evaluation, we added the following in the “2.7 validation and accuracy assessment” as below:

we selected 11 cities of varying sizes from across the study area for visual inspection in order to qualitatively evaluate the ability of U-Net to simulate realistic spatial urban land-use patterns and development characteristics.

We also discussed the reasons for using visual inspection to evaluate the performance of U-Net and discuss the need for better tools to quantitatively assess the complex spatial shapes in the urban land-use change process in the “4.5. Limitations and prospects” section:

Spatial patterns of urban development are highly complex. Very few tools currently exist to quantitatively compare these patterns. The shape index employed in this study revealed a high level of agreement between the U-Net projection of urban areas and the reference land-use map. However, our assessment of the ability of the U-Net to mimic specific patterns was largely based on visual inspection. No quantitative metrics currently exist to objectively assess whether projected urban patterns and shapes look plausible. Expert human interpretation such as the visual inspection method used in this study can effectively assess the realism of projection urban patterns but it is a qualitative and imprecise process. Better tools and metrics are required to quantitatively assess these complex spatial shapes and patterns to complement qualitative assessment based on human interpretation.

The highlight states "U-Net achieves similar accuracies with CA models with lengthen calibration". What do you mean about "lengthen calibration"? How does your study result support "similar accuracies"? What makes the U-Net models better than CA models? Perhaps, the highlight needs to be revised and a corresponding quantitative comparison and discussion are needed in the discussion section.

‘Lengthen calibration’ was a typo for which we apologise. With regards to comment on ‘similar accuracies’ we have completely rewritten this section (4.3. Accurate prediction and robustness in capturing spatial patterns) to improve the clarity of the comparison...

There are challenges to comparing model performance across studies. Pontius et al. (2008) reported two factors that profoundly influence urban land-use simulation accuracy: 1) the area of urban expansion, and; 2) the spatial resolution. A positive relationship exists between the FOM and observed land-use change. Prediction errors vanished when the simulation maps were resampled into coarser resolution (Pontius et al., 2008). Given these sensitivities, to enable a fair comparison of the accuracy of our U-Net model outputs with CA model outputs, we selected two CA-based studies with similar historical urbanization areas and the same 30m spatial resolution as our study. Wang et al. (2021c) used a particle swarm optimization algorithm, iterated a range of parameter settings such as a self-recognition component, a social component, inertia weights, and the number of particles, to finally arrive at the best model with an FOM of 0.193 for Zhuji, China. Wang et al. (2021a) developed 17 sub-models incorporating four periods of historical urban land-use and tested eight different neighborhood

sizes (5*5 to 41*41) for each sub-model to ultimately identify the best simulation which achieved a FoM of 0.219 for Beijing, China. The FoM (computed from 76 prefectures) in our study ranges from 0.177 - 0.215 (interquartile range) which covers the 0.193 reported by Wang et al. (2021c) but is slightly lower than the 0.219 found by Wang et al. (2021a). This higher FoM reported by Wang et al. (2021a) may be explained by the very large amount of urbanization in their study area (i.e., Beijing) (Pontius et al., 2008). **This comparison demonstrates that U-Net achieved similar predictive accuracies to comparable CA-based urban land-use models.**

One of my main concerns is the description of the data used in the study, which is not clear and consistent considering the following listed example lines. You actually used land use maps of 1994, 2006, and 2018 only, plus elevation/slope data (not sure if also of these three years), but the way the data are described is confusing.

We added table 3 in the “Data processing” section to make it clear to readers that only land-use maps (1994, 2006, and 2018) and elevation/slope data were used in this study.

Table 3. The data used to train U-Net models. (This table is inserted in the “2.4. Data preprocessing”)

Data	Time	Source	Resolution
Urban land-use	1994, 2006, 2018	Wang et al. (2021b)	30 m
Elevation	2000*	Shuttle Radar Topography Mission	30 m
Slope	2000*	Derived from elevation	30 m

* Note, the elevation and slope data were used in 1994, 2006, and 2018 despite their being acquired in 2000.

Line 97-98: "used high-accuracy maps of urban land-use between 1993 and 2012 to train the model" and tested the model by the projected spatial distribution of urbanization for 2018

We apologise for this error. We changed “between 1993 and 2012” to “between 1994 and 2006.”

To avoid confusion about how the U-Net was validated using a reference map, we modified the description to:

We then projected the spatial distribution of urbanization for 2018 using this model and thoroughly tested its ability to accurately capture the high-level spatial features, shapes, and patterns of urban development against the reference urban land-use of 2018.

Line 100-101: THIS trained model was then used to project patterns in 2030

To make it clear for readers which model was used to project the urban land-use in 2030, we modified the description as below:

Last, we trained another U-Net model on land-use maps for 2006 and 2018 to project urban land-use in 2030 based on the extrapolation of historical urbanization rates.

Line 111-113: here you start talking about using different years of data for two different U-net models.

In our initial submission, we denoted the different historical periods with different names (1994–2006 and 2006–2018 are referred to as “base,” 2018 and 2030 were called “target,” and the true 2018 urban map was called “reference”) and we described the U-Net as “U-Net trained on base/target years”. We now see that this notation could confuse the reader because they need to go back to the method section to discern the differences between “U-Nets trained on base/target years”. To avoid confusion between the two models trained in this study, we modified the description when introducing the two U-Net models in the “2.1. Method overview” section:

We created two U-Net models (Table 1) where one was trained/validated by historical data to evaluate the model performance and the other one was trained on more recent data to predict future urban land-use. U-Net-A was trained for the validation purpose by comparing its projected land-use map with the reference map using a range of accuracy and pattern-based metrics. U-Net-B was trained to predict future urban land-use for 2030.

Table 1. Brief description of the U-Net models trained in this study.

Model name	Training years	Prediction year
<i>U-Net-A</i>	<i>1994, 2006</i>	<i>2018</i>
<i>U-Net-B</i>	<i>2006, 2018</i>	<i>2030</i>

Line 108-109: "We used Landsat data to map urban development in the study area for the years 1994, 2006, and 2018" - I think you are talking about "map land use". The terms used throughout the manuscript need to be clearly defined or understood and consistent, e.g., urban dynamic map, urban map, urban land-use map, urban land use map, urban pixel number, urban pixels, urban images

Yes, this assumption is correct. We have now changed “urban dynamic map, urban map, etc.” to “urban land-use map” throughout the manuscript.

Line 196-199: what are the spatial resolutions of land-use, elevation, and slope data, assuming they are rasters?

We added a table 3 in the “Data processing” section to make it clear to readers that all data used were in raster format with a spatial resolution of 30m.

Line 213-215: what's urban map here? Here, what data were used to train which model?

We changed “urban map for 1994” to “urban land-use map for 1994.” To clarify the data used to train U-Net, we added more information in parentheses for explanation:

The samples included three-layer input data (e.g., urban land-use map for 1994, elevation and slope for U-Net-A) and a single-layer future urban land-use map (e.g., urban land-use map for 2006 for U-Net-A).

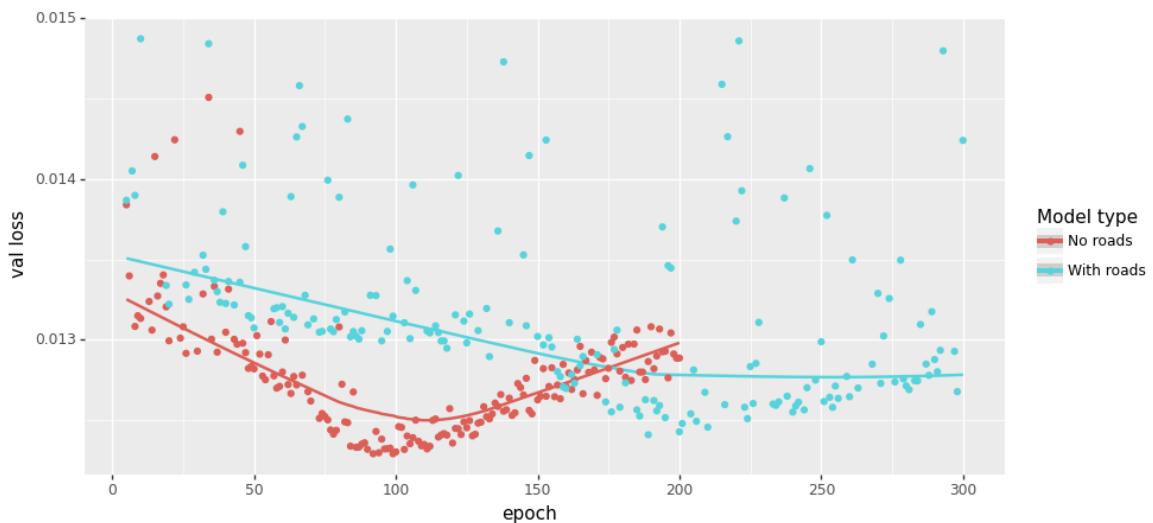
Line 226: what are the input images here? Land use maps?

The input images refer to the combination of the land-use map and the elevation/slope data. To clarify this concept, we added more information in the parenthesis for explanation:

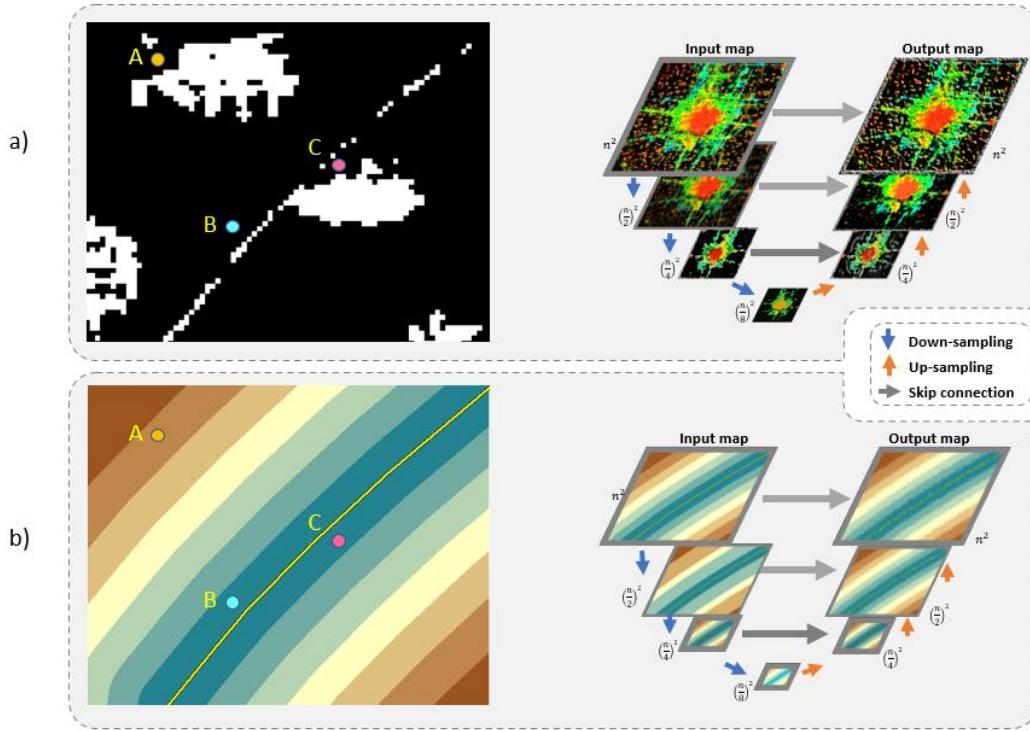
The input data (e.g., for U-Net-A a three-layer raster of urban land-use for 2006 and the elevation/slope data) was split into tiles, supplied to the trained U-Net model to produce separate outputs, and then the outputs were mosaicked into a single transition potential map.

Line 204-205: any previous studies that prove "little additional useful information" provided by distance factors?

To inform this decision, we conducted a separate experiment that showed that "distance to road provided little additional useful information." We collected historical road data (vector) from *OpenStreetMap*, computed the distance to roads in raster format, and appended this data as additional drivers to train the U-Net. The evaluation (MSE of 5k 256*256 tiled images against their corresponding reference 256*256 tiled images) shows that introducing road distance information makes the U-Net harder to train (longer training time and slightly higher MSE).



One possible explanation is that U-Net is sensitive to "visual signals (i.e., spatial features)" but distance hardly provides any robust spatial features. Below is a conceptual illustration of urban land-use (a) and distance-based driver (b) being processed by U-Net. The distance-based driver exhibits a monotonically increasing pattern perpendicular to the road, and this pattern remains unchanged in the U-Net. However, the overall urban land-use change patterns were captured in the down-sampling process and the finer patterns were reconstructed in the up-sampling process. Therefore, the monotonicity of distance-based drivers may prevent U-Net to learn additional patterns from roads/town centers, etc.



To make this clearer to readers, we provided a supplementary document to explain the reason for not including distance data in our study.

Line 285-286: "to historical urban land-use maps for the period 1994-2006" - the expression implies maps of more than two years

We changed this to "historical urban land-use maps of 1994 and 2006".

What message does Fig. 5 convey? How does it show the dynamics of urban land use? What do you mean about "urban dynamic map"? Do you really need this large map while showing the details as small snapshots?

The purpose of Fig. 5 was to introduce the historical urban land-use in the study area to the reader. We agree that this figure is probably extraneous and have now moved it to supplementary documents.

Line 310-312: Is SuiXian a large town? What are those smaller villages which have the same phenomena - high transition potential?

There are several small towns, including Suixian, which have been selected to illustrate transition potential calculated by U-Net. The high transition potential effects are because of a "standard deviation" stretch method being applied to the map. We have now realized that applying a stretch method would exaggerate the transition potentials surrounding initial urban area and removed the stretching method.

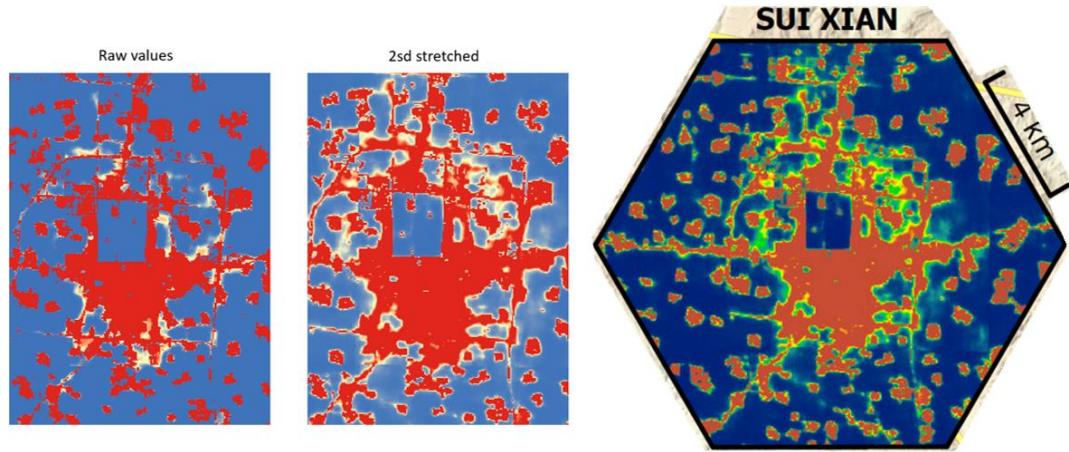


Figure 10 - do you really need this large map while showing the details as small snapshots?

Figure 11 - do you really need this large map while showing the details as small snapshots?

We do need these two figures to demonstrate that U-Net had learned the “gravity, neighborhood, and linear effects” in urban development via visual inspection. We chose to keep the large map as well as show the detailed inset maps to demonstrate to readers the overall view of the whole study area on how U-Net performed and the degree of urban expansion. Also, we found that eliminating the large map won’t necessarily reduce much page space (see below our alternative design) and readers will wonder where these localities are within the region.

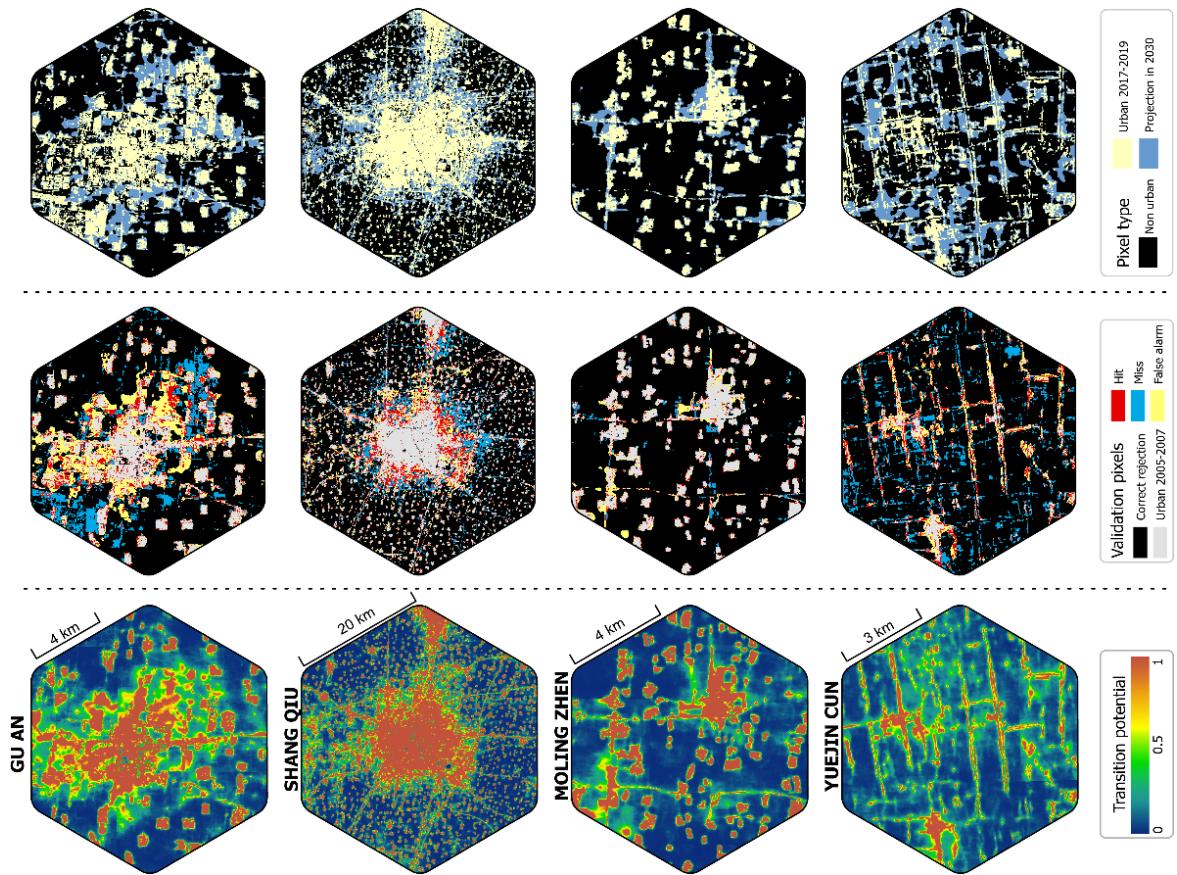
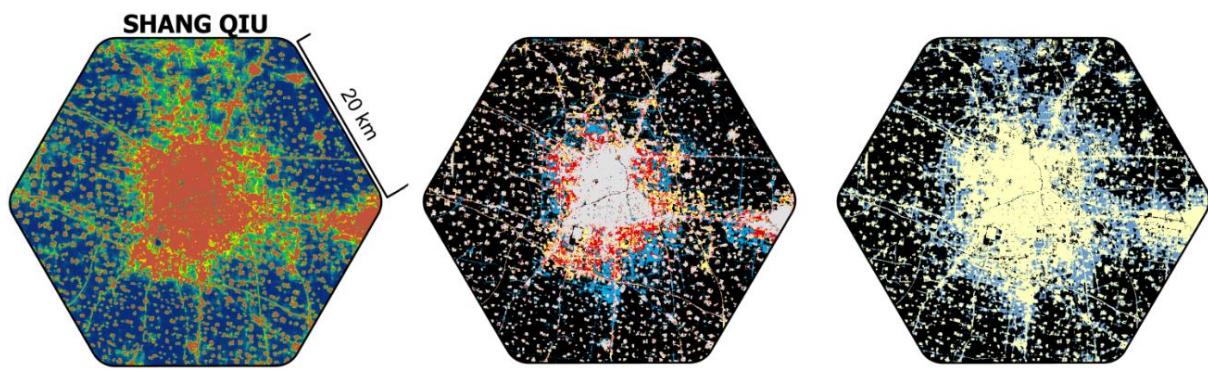


Fig. 10, 11, and 13 show 11 smaller urban areas, no large urban centers. How will you generalize the observations from the patterns around these areas to more general urban development?

Readers can zoom into the large cities to see the generalized simulation of urban land-use with our high-resolution figure. The data will also be made available freely online. We have tried to include large cities as inset maps, but the visual effects are less clear than in the small towns because the subtle details are obscured by large-scale transition potential patterns (see below figure).



What's the purpose of including the projection of 2030 urban land use? Is it just to demonstrate the use of the trained U-Net models?

The complete workflow of urban land-use simulation in this study is: 1) develop the U-Net architecture, 2) validate the trained U-Net against historical reference urban land-use map, and 3) apply the U-Net to project urban land-use to a future date. As the reviewer rightly recognizes, the primary purpose of including the projection to 2030 is to demonstrate the use of the trained U-Net. However, as one of the fastest urbanized regions in the world as well as the food bowl of China, knowing the accurate future urban land-use in the future provides significant practical value to support the regulation of urban development and evaluation of the impacts on sustainability. Urban land-use is a key proxy for socioeconomic assessment and environmental protection. Studies focusing on sustainable development and climate change can benefit from our accurate future land-use predictions.

We explained the importance of future urban land-use to North China plain in the 2.2. Study area section and future discussed the implication of the prediction in the discussion section.

This region is crucial to China's economic development and holds a strategic role in safeguarding China's food security, generating over one-third of its national gross domestic product (GDP) and grain supply (National Bureau of Statistics of China, 2019a). Managing the tension between urbanization and agricultural land-use in the study area requires accurate, spatially explicit projections of future urban development to address the interconnected challenges of food security, environmental protection, urbanization, and socio-economic development.

... Our projections for 2030 also captured urban development in the three megacity groups (Beijing-Tianjin-Hebei, the Yangtze River Delta, and the Central Plains) that account for one-third of China's GDP (National Bureau of Statistics of China, 2019b). China's strategic development planning process can benefit from the predictions arising out of this study in many ways. For example, in urban area predictions can be used to plan for infrastructure to support socio-economic development. Simulations can assist policy formulation that is tailored to the expected rate, location, and patterns of urbanization ...

Reviewer #2:

I have reviewed this article carefully; I agree that it is a promising new idea to apply U-Net to urban growth simulation and prediction.

Thanks for confirming the value of this study.

However, this method focuses on the characteristics of images (i.e., urban land use landscape), rather than understanding the process of urban growth from the driving factors (i.e., the combined effects of nature and society, etc.) that drive landscape evolution. Therefore, this method cannot understand the mechanism of urban evolution.

There are two types of CA-based urban land-use simulation models: one is called a regressive model which focuses on increasing the simulation accuracy and finding the optimum functions to decompose drivers (e.g., suitability, proximity, stochastic factors, etc.) into transition rules (Gantumur et al., 2020; Shafizadeh-Moghadam et al., 2017). The other one, often called the participatory model, emphasizes balancing the interests of stakeholders in the urbanization process to reveal the mechanisms/drivers behind urban land-use change (Clarke and Johnson, 2020; Mansour et al., 2020).

The U-Net shares the same goal the regressive model of increasing prediction accuracies but does so using a deep learning approach rather than building a regression between driving factors and urban land-use. It possesses several significant advantages: 1) it automatically extracts transition rules without the requirement of manually setting parameters such as neighborhood size and decay rates (Feng and Tong, 2020; Peng et al., 2020; Roodposhti et al., 2020), and 2) it uses computer vision technologies to learn the spatial patterns of urban development. Therefore, deep learning architectures bring about the significant potential to complement CA-based models with comparable simulation performances and the automatic recognition of complex spatial patterns.

The U-Net model revealed, via visual inspection, three significant urban development processes such as neighborhood effects, gravity effects, and linear expansion effects. However, these learned patterns are hidden in the ~30 million weights within the U-Net. There are several potential ways to transform the trained U-Net into human recognizable rules, i.e., revealing the mechanism of urban evolution. The explainable machine learning techniques enable the abstracting of knowledge from deep learning architectures (Dosilovic et al.; Gunning et al., 2019). However, there are uncertainties if these measures can be successfully applied to U-Net and they also exceeded the topic of this study. To make readers clear about the limited capability of U-Net models as well as the potential ways to extract human-recognizable rules from them, we described the advancement in the deep learning technology as a potential solution to help understand the urban land-use change mechanism in the discussion section:

Emerging developments in extracting human-recognizable knowledge from deep learning architectures (Dosilovic et al.; Gunning et al., 2019) could be applied to U-Net models for a better understanding of the mechanisms behind urban land-use dynamics.

On the other hand, the author does not strongly prove that the simulation accuracy of this method is better than the results of traditional CA-based results. I noticed that the author mentioned in the article that U-NET achieved higher accuracy than CA in some cities. But this does not mean that U-NET performs better in a large area, such as the study area of this paper.

The primary goal of introducing deep learning to urban land-use simulation is to compare the pros and cons and explore how deep learning could complement CA-based land-use modeling, rather than to demonstrate that deep learning is superior. The comparison between U-Net and other studies showed that these two methods are very comparable from an accuracy metrics perspective.

The accuracy of the simulated urban land-use is profoundly impacted by 1) the net urbanized area and 2) the spatial resolution of the raster data (Pontius et al., 2008). We compared our result with two CA-based studies that iterated multiple parameters (window sizes, inertia weights, iteration number, etc.) or developed 17 sub-models to get the best simulation outcome (Wang et al., 2021c; Wang et al., 2021a). The comparison showed that the FoM in this study, computed from 76 prefectures, with the 1st quantile value of 0.177, the 3rd quantile of 0.215, and the median of 0.197, was close to 0.193 reported by Wang et al. (2021c), and 0.219 found by Wang et al. (2021a).

U-NET achieved higher accuracy than CA in some cities because the FoM is positively related to the urbanization rate (Pontius et al., 2008): some cities had more net urbanized areas and are more likely to have higher FoM. However, the interquartile value of FoM from U-Net simulation is very close to the selected CA-based studies.

We have modified the manuscript where comparing U-Net with CA-models as bellow:

*Wang et al. (2021c) used a particle swarm optimization algorithm, iterated a range of parameter settings such as a self-recognition component, a social component, inertia weights, and the number of particles, to finally arrive at the best model with an FoM of 0.193 for Zhuji, China. Wang et al. (2021a) developed 17 sub-models incorporating four periods of historical urban land-use and tested eight different neighborhood sizes (5*5 to 41*41) for each sub-model to ultimately identify the best simulation which achieved an FoM of 0.219 for Beijing, China. The FoM (computed from 76 prefectures) in our study ranges from 0.177 - 0.215 (interquartile range) which covers the 0.193 reported by Wang et al. (2021c) but is slightly lower than the 0.219 found by Wang et al. (2021a). This higher FoM reported by Wang et al. (2021a) may be explained by the very large amount of urbanization in their study area (i.e., Beijing) (Pontius et al., 2008). This comparison demonstrates that U-Net achieved similar predictive accuracies to comparable CA-based urban land-use models.*

References

- Chen, S., Feng, Y., Tong, X., Liu, S., Xie, H., Gao, C., & Lei, Z. (2020). Modeling ESV losses caused by urban expansion using cellular automata and geographically weighted regression. *The Science of the total environment*, 712, 136509.
- Clarke, K.C., & Johnson, J.M. (2020). Calibrating SLEUTH with big data: Projecting California's land use to 2100. *Computers, Environment and Urban Systems*, 83, 101525.
- Dosilovic, F.K., Brcic, M., & Hlupic, N. Explainable artificial intelligence: A survey, 210–215.
- Feng, Y., & Tong, X. (2020). A new cellular automata framework of urban growth modeling by incorporating statistical and heuristic methods. *International Journal of Geographical Information Science*, 34, 74–97.

- Frazier, A.E., & Kedron, P. (2017). Landscape Metrics: Past Progress and Future Directions. *Current Landscape Ecology Reports*, 2, 63–72.
- Gantumur, B., Wu, F., Vandansambuu, B., Tsegmid, B., Dalaibaatar, E., & Zhao, Y. (2020). Spatiotemporal dynamics of urban expansion and its simulation using CA-ANN model in Ulaanbaatar, Mongolia. *Geocarto International*, 1–16.
- Ghosh, B., Dutta, I.K., Totaro, M., & Bayoumi, M. (2020). A Survey on the Progression and Performance of Generative Adversarial Networks, 1–8.
- Gonog, L., & Zhou, Y. (2019). A Review: Generative Adversarial Networks, 505–510.
- Gunning, D., Stefk, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI-Explainable artificial intelligence. *Science robotics*, 4.
- Mansour, S., Al-Belushi, M., & Al-Awadhi, T. (2020). Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques. *Land use policy*, 91, 104414.
- Mustafa, A., Cools, M., Saadi, I., & Teller, J. (2017). Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM). *Land use policy*, 69, 529–540.
- National Bureau of Statistics of China (2019a). Announcement of the 2019 grain output.
http://www.gov.cn/xinwen/2019-12/07/content_5459250.htm.
- National Bureau of Statistics of China (2019b). *China Statistical Yearbook*. Beijing, China: China Statistics Press.
- Peng, K., Jiang, W., Deng, Y., Liu, Y., Wu, Z., & Chen, Z. (2020). Simulating wetland changes under different scenarios based on integrating the random forest and CLUE-S models: A case study of Wuhan Urban Agglomeration. *Ecological Indicators*, 117, 106671.
- Pontius, R.G., Boersma, W., Castella, J.-C., Clarke, K., Nijs, T. de, Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T.N., Veldkamp, A.T., & Verburg, P.H. (2008). Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42, 11–37.
- Roodposhti, M.S., Hewitt, R.J., & Bryan, B.A. (2020). Towards automatic calibration of neighbourhood influence in cellular automata land-use models. *Computers, Environment and Urban Systems*, 79, 101416.
- Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., & Taleai, M. (2017). Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, 64, 297–308.
- Tripathy, P., & Kumar, A. (2019). Monitoring and modelling spatio-temporal urban growth of Delhi using Cellular Automata and geoinformatics. *Cities*, 90, 52–63.
- Uuemaa, E., Antrop, M., Roosaare, J., Marja, R., & Mander, Ü. (2009). Landscape metrics and indices : an overview of their use in landscape research. *LIVING REVIEWS IN LANDSCAPE RESEARCH*, 3.
<http://biblio.ugent.be/publication/695518>.
- Wang, H., Guo, J., Zhang, B., & Zeng, H. (2021a). Simulating urban land growth by incorporating historical information into a cellular automata model. *Landscape and Urban Planning*, 214, 104168.
- Wang, J., Hadjikakou, M., & Bryan, B.A. (2021b). Consistent, accurate, high resolution, long time-series mapping of built-up land in the North China Plain. *GIScience & Remote Sensing*, 58, 982–998.
- Wang, R., Feng, Y., Wei, Y., Tong, X., Zhai, S., Zhou, Y., & Wu, P. (2021c). A comparison of proximity and accessibility drivers in simulating dynamic urban growth. *Transactions in GIS*, 25, 923–947.
- Wang, Z., Healy, G., Smeaton, A.F., & Ward, T.E. (2020). Use of Neural Signals to Evaluate the Quality of Generative Adversarial Network Performance in Facial Image Generation. *Cognitive Computation*, 12, 13–24.
- Xu, Q., Huang, G., Yuan, Y., Guo, C., Sun, Y., Wu, F., & Weinberger, K. (2018). An empirical study on evaluation metrics of generative adversarial networks.

Yu, J., Hagen-Zanker, A., Santitissadeekorn, N., & Hughes, S. (2021). Calibration of cellular automata urban growth models from urban genesis onwards - a novel application of Markov chain Monte Carlo approximate Bayesian computation. *Computers, Environment and Urban Systems*, 90, 101689.

Highlights

- U-Net deep learning architecture was able to simulate complex urban development patterns
- Urban transition rules were learned automatically in the U-Net with minimal data requirements
- The complex spatial patterns of simulated urban land-use closely matched actual patterns
- U-Net achieved similar accuracies to CA-based models with a comparable urbanization context
- U-Net learned the neighborhood, gravity, and linear effects of urban growth

Simulating large-scale urban land-use patterns and dynamics using the U-Net deep learning architecture

Jinzhu Wang^{a,b*}, Michalis Hadjikakou^a, Richard J.Hewitt^{c,d,e}, Brett A. Bryan^{a,b}

^a Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, VIC 3125, Melbourne, Australia.

^b Deakin-SWU Joint Research Centre on Big Data, Faculty of Science, Engineering and Built Environment, Deakin University, VIC 3125, Australia.

^c Transport, Infrastructure, and Territory Research Group (*tGIS*), Geography Department, Faculty of Geography and History, Universidad Complutense de Madrid (UCM), C/ Profesor Aranguren, s/n, Ciudad Universitaria, 28040, Madrid, Spain.

^d Observatorio para una Cultura del Territorio (OCT), Calle del Duque de Fernán Núñez 2, 1, 28012, Madrid, Spain.

^e Informational and Computational Sciences Group, The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland UK.

*corresponding author at wangjinz@deakin.edu.au; Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, VIC 3125, Melbourne, Australia;

Abstract

Accurate predictions of land-use change are important for supporting planning. Cellular automata (CA) models are widely used to simulate real-world urban land-use change but accurately modelling complex spatial urban patterns and dynamics can be challenging due to the high degree of subjectivity involved in CA model parameterisation. Advances in deep learning enable complex spatial patterns such as urban development to be learned and simulated. In this study, we used the U-Net deep learning algorithm to capture historical urban development and simulate future patterns for the North China Plain, one of the most rapidly urbanizing regions on the planet. We validated the model against a reference map for 2018 then applied it in predicting patterns of urban expansion for 2030. The results showed that U-Net can accurately predict urban land-use and mimic real-world spatial patterns with very low requirement for model parameterization and forcing data. Visual inspection of the outputs revealed that U-Net was able to automatically learn complex urban development patterns and processes such as neighbourhood influence, the gravity effects of large cities, and the tendency for linear development. Deep learning architectures such as U-Net provide a new parameter-free way to accurately capture and simulate spatial features in projections of future urban development and land-use change.

Keywords: Deep Learning; urban expansion simulation; cellular automata; spatial pattern

Acknowledgment

A Deakin University Postgraduate Research scholarship funded this research.

1 1. Introduction

2 Urbanization is a complex process that is influenced by a range of social, cultural, economic,
3 geographic, environmental, and political factors (Shaw et al., 2020; Yeh and Chen, 2020; Kipfer,
4 2018; Fan et al., 2018). Understanding future urban patterns and their dynamics is essential for
5 ensuring sustainable development due to the profound environmental and socio-economic impacts
6 of urban land transformation (Zheng et al., 2017; Newland et al., 2018). Modeling urban patterns
7 and dynamics has been undertaken in many parts of the world as a basis for mitigating air pollution
8 (Fan et al., 2020), habitat destruction (Planillo et al., 2021), and loss of arable land (Qiu et al., 2020).
9 Some urban modeling studies have focused on participatory modeling and scenario analysis for
10 engaging stakeholders and experts in the modeling process to balance competing interests and
11 facilitate co-learning processes related to future urban development (Clarke and Johnson, 2020;
12 Mansour et al., 2020; Peng et al., 2020). However, most urban land-use modeling studies have
13 focused primarily on the accurate prediction of the future spatial layout of cities based on historical
14 dynamics and changes in key driving forces under various scenarios (Gantumur et al., 2020;
15 Shafizadeh-Moghadam et al., 2017). Despite significant recent advances in urban land-use modeling,
16 accurately capturing the complex spatial dynamics, patterns, and stochasticities of cities remains a
17 significant challenge (Liu et al., 2021).

18 Cellular automata (CA) have been widely used to model future urbanization patterns because of
19 their ability to incorporate real-world urbanization processes (Tong and Feng, 2020; Li et al., 2017).
20 CA models consider multiple factors such as land suitability, neighborhood status, constraint
21 variables, and stochastic factors (Roodposhti et al., 2020; Wang et al., 2021a). Land suitability refers
22 to the rescaled biophysical, geographic, and socio-economic driving factors that influence
23 urbanization (Feng and Tong, 2020). Neighborhood status reflects the amount of urban or other
24 land-use occurring in the immediate vicinity of each cell and is characterized by different structures,
25 sizes, and weights (Yu et al., 2021; Roodposhti et al., 2020). Constraint variables and stochastic
26 factors regulate and randomize future urban development, respectively (Zhai et al., 2020). Transition
27 rules are a set of functions and parameters that control the scaling of land suitability factors,
28 configuration of the neighborhood, constraints, and stochastic factors, and then combine these
29 elements into a spatial layer defining the probability of each cell becoming urbanized in the future.
30 While transition rules have typically been derived by trial and error or expert knowledge, they are
31 increasingly derived automatically to achieve the highest predictive accuracy. Automatic rule
32 extraction has included a suite of regression and machine learning (ML) based methods such as
33 logistic regression (Mustafa et al., 2018), support vector machines (Kafy et al., 2021), tree-based
34 methods (Shafizadeh-Moghadam et al., 2017), neural networks (Gantumur et al., 2020), heuristic
35 methods (Carneiro and Oliveira, 2013), and dictionaries of trusted rules (Roodposhti et al., 2019).
36 Although the flexibility of process-based CA-based models with their large array of parameter
37 settings makes them ideal for participatory scenario modeling exercises, the difficulty in calibrating
38 the many parameter choices—still largely a manual process of trial and error—challenges their
39 ability to mimic complex urban land-use and accurately capture future urban patterns.

40 More recent studies have adopted different techniques, such as geographical zoning, context
41 integration, and innovative algorithms, to increase the predictive accuracy of urban land-use

42 modeling. For example, some studies have subdivided their study area into separate regions,
43 allowing independent transition rule sets to be constructed to align with specific conditions in each
44 zone (Qian et al., 2020; Xia and Zhang, 2021). Many studies have incorporated the shape and texture
45 index to reflect the neighborhood spatial configurations of urban land-use (Zhai et al., 2020; Ruiz
46 Hernandez and Shi, 2018). Wang et al. (2021a) incorporated historical urban development as a
47 temporal process to simulate future urban land-use, while Peng et al. (2020) integrated evolutionary
48 and swarm algorithms to mimic urban land-use. Despite having successfully constructed transition
49 rules, these studies identified significant mismatches when compared to real-world urbanization.
50 The source of these errors has been identified as: 1) the introduction of subjectivities via prescribed
51 parameters (Liu et al., 2021), and; 2) oversimplifying the spatial heterogeneity of the driving factors
52 (Gao et al., 2020; Mustafa et al., 2018; Newland et al., 2018). Distances to geographical factors and
53 the spatial configurations of these factors are key to urban development in the real world. The
54 distances used in these studies however, are measured by preset decay functions, and subjectivities
55 are inevitably raised, given the gap between empirical knowledge and real-world urban land-use.
56 The spatial heterogeneity of the urbanization process is often oversimplified to a single shape or
57 texture index that captures only part of the pattern information within the neighborhood (Akın and
58 Erdoğan, 2020; Motlagh et al., 2020), let alone the large-scale spatial features (e.g., the layout of the
59 whole built city area) that typify real world urban development.

60 Deep learning algorithms are a type of machine learning technology that abstract naïve input
61 variables into high-level features. Convolutional neural networks (CNNs) are a special kind of deep
62 learning architecture designed to detect spatial patterns from both low- (e.g., lines, triangles, circles)
63 and high-level features (e.g., human faces, animals) (Krizhevsky et al., 2017). CNNs are increasingly
64 being used to extract patterns and insights from geospatial data (Reichstein et al., 2019) which
65 enables the spatial configurations of driving factors, rather than proxy variables such as decay
66 distances, to be integrated directly into urban development simulations. For example, Zhai et al.
67 (2020) used a CNN to retrieve the neighborhood spatial features and improve the Figure of Merit
68 (FoM) of their simulation to 0.361 compared to a random-forest-based method (0.323). Qian et al.
69 (2020) reported similar results, i.e., improving the FoM from 0.299 to 0.346 compared to a random-
70 forest algorithm. Although the spatial features introduced by the CNN improved the simulation
71 performance, these studies used CNNs as little more than an advanced decay function to rescale the
72 driving factors and overlooked their ability to abstract high-level features. More advanced deep
73 learning structures that take full advantage of the spatial pattern recognition ability of CNNs, i.e.,
74 integrating low-level spatial features into high-level patterns, have great potential for the spatial
75 simulation of future urban development.

76 U-Net, first introduced for biomedical image segmentation in 2015, is a unique type of CNN
77 architecture that not only abstracts spatial features to high-level patterns but also refines the high-
78 level patterns to precise shapes (Ronneberger et al., 2015). The robust segmentation performance of
79 U-Net enables it to be used in multiple fields such as improving weather prediction (Singh et al.,
80 2021), detecting underwater objects for oceanic ecosystem evaluation (Nezla et al., 2021), and
81 identifying buildings from aerial and satellite imagery (Ji et al., 2019). Unlike CA models that have a
82 fixed neighborhood size, U-Net deploys a series of convolutional layers to extract spatial features
83 and then assimilates these features automatically to produce transition rules. This ability to learn

84 spatial patterns suggests that U-Net has the potential to identify and assimilate the spatial processes
1 85 that drive urban development and accurately capture the resulting patterns of cities.
2

3
4 86 In this study, we used the U-Net deep learning architecture to project urban land-use in the North
5 87 China Plain—China's food bowl and one of the most rapidly urbanizing areas on Earth. We first
6 88 applied a U-Net model to learn the patterns of urban land-use and land-use change between 1994
7 89 and 2006. We then projected the spatial distribution of urbanization for 2018 using this model and
8 90 thoroughly tested and validated its ability to accurately capture the high-level spatial features,
9 91 shapes, and patterns of urban development against a reference land-use map for 2018. Last, we
10 92 trained another U-Net model on land-use maps for 2006 and 2018 to project urban land-use in 2030
11 93 based on the extrapolation of historical urbanization rates. We discuss the advantages and
12 94 limitations of U-Net for simulating urban development and, in particular, its ability to learn urban
13 95 land-use patterns such as neighborhood influences and linear expansion along transport routes. We
14 96 also discuss the implications of future rapid urbanization on sustainable development in the study
15 97 area.
16
17

18 98

2. Methods

19

20 99

2.1. Method overview

21

22 100 We used Landsat data to map urban land-use in the study area for the years 1994, 2006, and 2018
23 101 (Wang et al., 2021b) and combined them with elevation and slope information to simulate urban
24 102 development. We created two U-Net models (Table 1) where one was trained/validated by historical
25 103 data to evaluate the model performance and the other one was trained on more recent data to
26 104 predict future urban land-use. *U-Net-A* was trained for the purpose of validation by comparing its
27 105 projected land-use map with the reference map using a range of accuracy and pattern-based
28 106 metrics. *U-Net-B* was trained to predict future urban land-use for 2030. The training samples for
29 107 both U-Net models were randomly extracted from the data for the training years. The trained
30 108 models were then applied to produce a transition potential layer used to create an urban land-use
31 109 map for the prediction year. The study workflow is summarized and illustrated in Fig. 1 and
32 110 described in more detail below.
33
34

35 111 Table 1. Brief description of the U-Net models trained in this study.
36 112
37

44 Model name	45 Training years	46 Prediction year
47 U-Net-A	48 1994, 2006	49 2018
50 U-Net-B	51 2006, 2018	52 2030

113

114

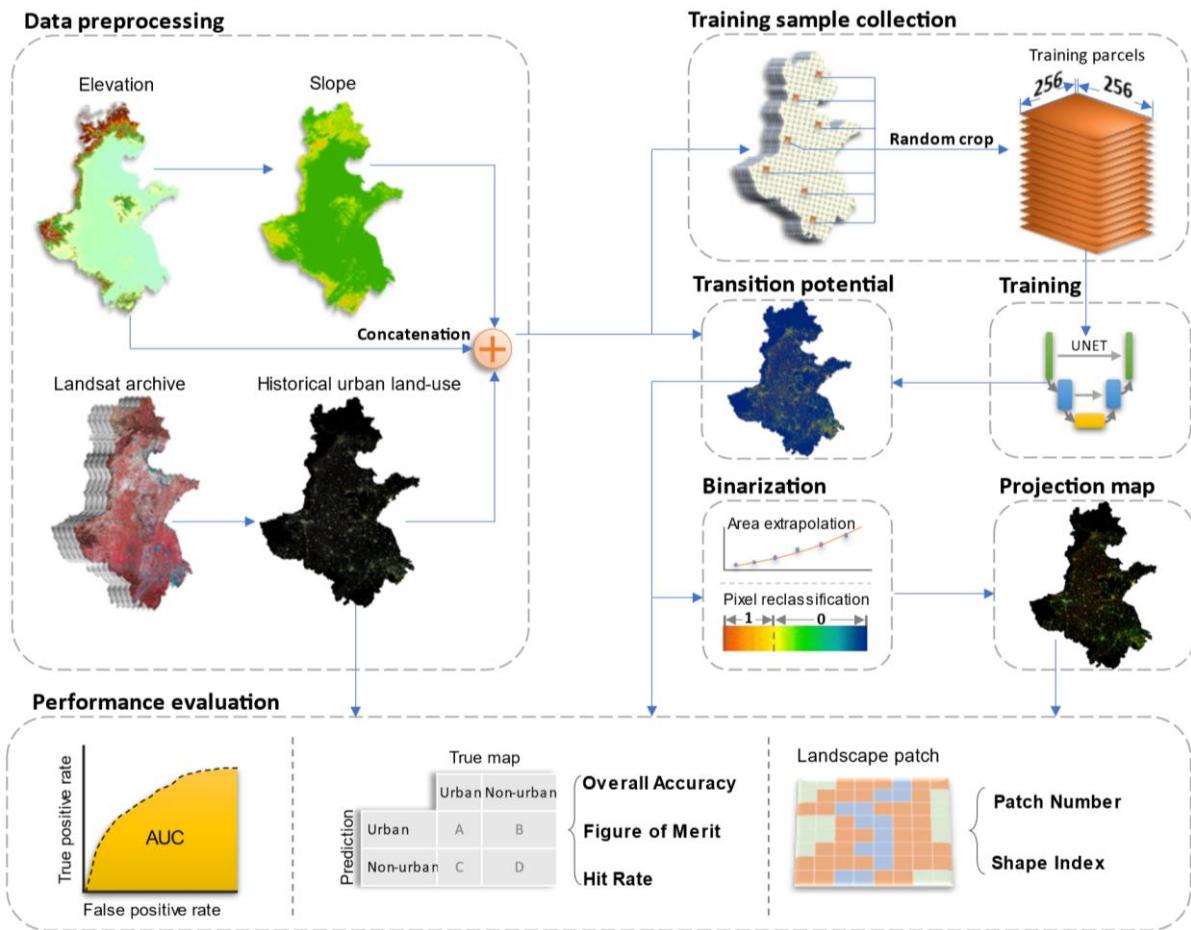


Fig. 1. Research workflow.

115

116

2.2. Study area

The North China Plain (Fig. 2) includes 76 prefectures, spans an area of more than 780,000 km², and is home to more than 450 million people (National Bureau of Statistics of China, 2019b). This area is one of the most rapidly urbanizing regions in China and the world, tripling its built-up land coverage from approximately 5% in 1990 to approximately 15% in 2020 (Wang et al., 2021b). This region is crucial to China's economic development and holds a strategic role in safeguarding China's food security, generating over one-third of its national gross domestic product (GDP) and grain supply (National Bureau of Statistics of China, 2019a). Managing the tension between urbanization and agricultural land-use in the study area requires accurate, spatially explicit projections of future urban development to address the interconnected challenges of food security, environmental protection, urbanization, and socio-economic development.



Fig. 2. Study area of the North China Plain.

2.3. Structure of U-Net

The U-Net structure includes down-sampling layers which extract the general context from the input data, up-sampling layers which refine these contexts to precise shapes, and skip-connections which balance the generalization of down-sampling and the refinement of up-sampling (Ronneberger et al., 2015). A conceptual U-Net structure (Fig. 3) demonstrates its pattern recognition capability.

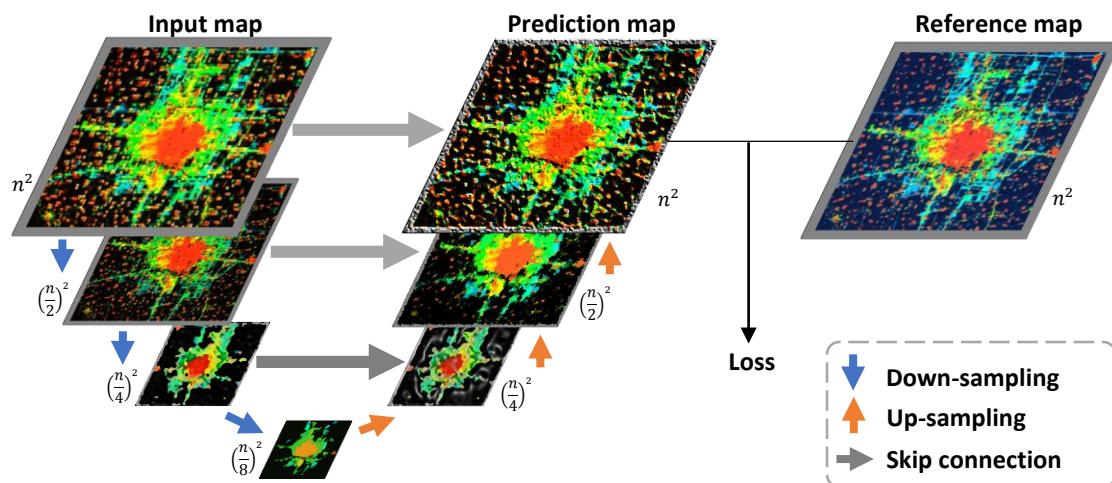


Fig. 3. Conceptual structure of a four-layer U-Net. In this study, the input map was resized to half of its input size (e.g., n^2 to $(\frac{n}{2})^2$) in each down-sampling layer, and then expanded to the original size in the up-sampling process. The loss denotes the difference between the prediction and reference maps, reflecting the performance of the U-Net.

139 The down-sampling layers include convolution, pooling, and rescaling processes (Table 2).
 140 Convolution is a cross-correlation operation that produces feature maps indicating the similarities
 141 between inputs and convolution filters (Kapinchev et al., 2015). By applying multiple filters, the
 142 different spatial patterns can be retrieved independently. For example, Zeiler and Fergus (2013)
 143 reported that horizontal, vertical, and circular patterns were identified by different filters applied to
 144 the input image. The pooling process reduces the size of inputs. We selected max-pooling to reduce
 145 dimensionality because it is adopted in most deep learning structures (Murray and Perronnin, 2014).
 146 The rescaling process rescales the pixel values of the feature maps to a specified range to optimize
 147 the computational flow of the network. We used a rectified linear unit activation function (ReLU) for
 148 its simple, efficient, and robust performance (Agarap, 2018).

149 Table 2. Equations of the component layers of U-Net.

Layer	Equation	No.
Convolution	$C_{\text{out},j} = \text{bias}(C_{\text{out}}) + \sum_{k=0}^{C_{\text{in}}-1} \text{weight}(C_{\text{out},j}, k) * \text{input}(k)$	(1)
Max-pooling	$\text{out}(C_j, h, w) = \max_{m=0, \dots, kH-1} \max_{n=0, \dots, kW-1} \text{input}(C_j, \text{stride}[0] \times h + m, \text{stride}[1] \times w + n)$	(2)
ReLU	$\text{ReLU}(x) = \max(0, x)$	(3)
Softmax	$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^K \exp(x_j)}$	(4)
Cross-entropy	$\text{loss}(x, y) = -\log \left(\frac{\exp(x[y])}{\sum_{j=1}^K \exp(x[j])} \right)$	(5)

37
 38 In equation (1), the sizes of the input and output images are (C_{in}, H, W) and $(C_{\text{out}}, H_{\text{out}}, W_{\text{out}})$, C denotes the number of
 39 channels, H is the height of the input planes in pixels, W is the width in pixels, $*$ is the valid cross-correlation operator, and j
 40 is the j -th channel of the output feature map. In equation (2), (kH, kW) denotes the kernel size of the pooling, h and w refer
 41 to the height and width of the output image, respectively. In equation (3), x denotes the pixel values of the input feature
 42 map. In equation (4), x_i is the i -th pixel value of the input feature map and K is the number of classes. In equation (5), x and
 43 y refer to the predicted and reference pixel values, respectively, and K is the number of classes.
 44

45 The up-sampling layers include transpose convolution and rescaling processes. The transpose
 46 convolution, an inverse convulsive operation, transforms the input image from a lower resolution
 47 to a higher one. This process is assisted by skip-connections that bring additional spatial information
 48 from the down-sampling layers. The transpose convolution is similar to the convolution except that
 49 the output size is larger than the input size and the rescaling is the same ReLU operation in the
 50 down-sampling. Additional components used in the U-Net are batch normalization, softmax, and the
 51 cross-entropy algorithm (Table 2). Batch normalization is applied to standardize the weights that
 52 control the convolution and transpose convolution processes, which has been proven to be effective
 53 in improving deep learning performance (Ioffe and Szegedy, 2015). The softmax algorithm is applied
 54 to the last feature map of the U-Net to rescale the pixel values to the 0–1 range for a better
 55 comparison with the reference map (also composed of pixels of 0s and 1s). The cross-entropy
 56

167 algorithm is then applied to calculate the difference between the prediction and the reference map
 168 and reflects the performance of the U-Net.

169 The complete U-Net structure used in this study is shown in Fig. 4. The input image size was reduced
 170 by half after passing each down-sampling block until reaching a size of 8×8 pixels (i.e., 8^2 in Fig. 4),
 171 then the image was ultimately recovered to the original size through each up-sampling block. The
 172 channels of the feature maps underwent the inverse process: the number of feature maps was
 173 doubled in each down-sampling block and halved in each up-sampling block. As a result, the U-Net
 174 model gradually extracted more abstract spatial features over a larger field of view according to $k \times$
 175 $2^{(d-1)}$, where k is the kernel size (3 in this study) and d is the depth of the layer block. For example,
 176 the block with depth 1 had a field of view of 3, which became 96 in the block of depth 6, meaning
 177 that U-Net was looking for spatial patterns of 96×96 pixels in the original image scale. Meanwhile,
 178 U-Net identified more sophisticated spatial patterns as the network block went deeper because the
 179 number of feature maps was doubled. The skip-connection links the down-sampling and up-
 180 sampling blocks, allowing U-Net to refine the overview patterns extracted by deeper network blocks
 181 with precise shapes and textures retrieved by shallower blocks. A total of more than 31 million
 182 parameters were included in the U-Net model, making it highly flexible for capturing the spatial
 183 patterns and stochasticity of the urban land-use process.

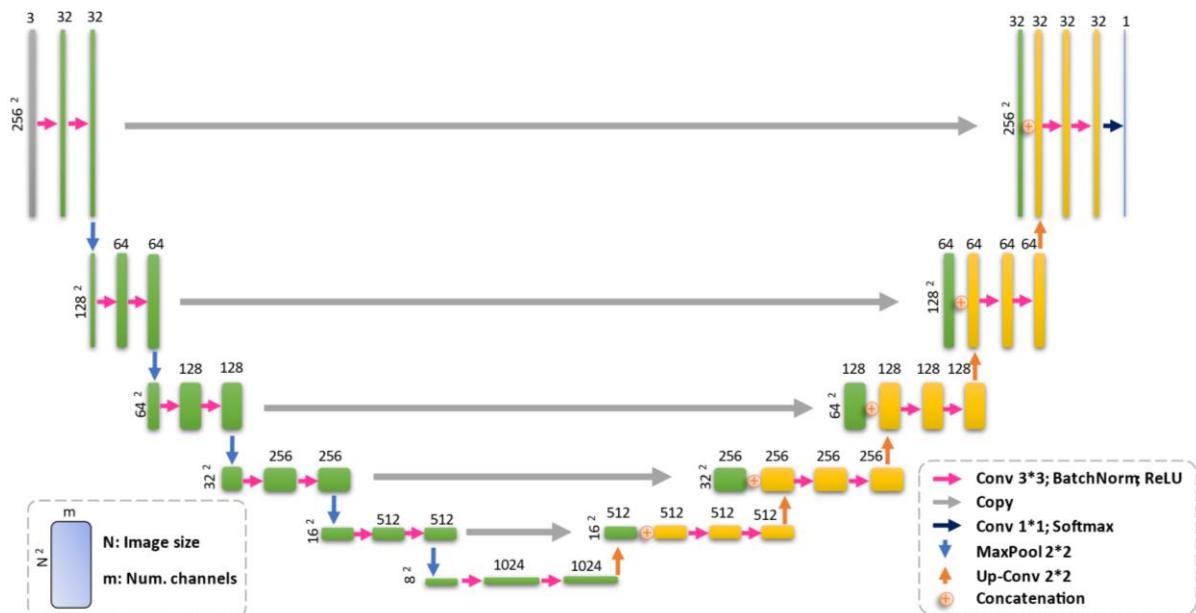


Fig. 4. Structure of the U-Net models used in this study.

2.4. Data preprocessing

We mapped urban land-use for the years 1994, 2006, and 2018 with a consistently high (>94%) accuracy (for details see Wang et al. (2021b)). Terrain data was obtained from the Shuttle Radar Topography Mission, from which slope and elevation data were derived (Table 3). Elevation and slope information was used to assist U-Net in capturing the topographic control of urban development (i.e., urban expansion is more likely on flatter vs hilly and mountainous landscapes), a strategy that has proven to be effective in previous studies (Xing et al., 2020; Wang et al., 2021a; Qian et al., 2020). While accessibility variables such as distance to roads and railways are frequently employed in urban land-use modeling (Tripathy and Kumar, 2019; Valencia et al., 2020; Ronneberger

et al., 2015), U-Net is a pattern-sensitive model and in this case, these data layers provide little additional useful information (see Supplementary Material). Hence, to maintain model parsimony and reduce training times distance variables were not used in this study.

Table 3. The data used to train U-Net models.

Data	Acquiring Time	Source	Resolution
Urban land-use	1994, 2006, 2018	Wang et al. (2021b)	30 m
Elevation	2000*	Shuttle Radar Topography Mission	30 m
Slope	2000*	Derived from elevation	30 m

* Note, the elevation and slope data were used in 1994, 2006, and 2018 despite their being acquired in 2000.

2.5. Training the U-Net models

Control samples were assembled using Google Earth Engine (Gorelick et al., 2017). The *NeighborhoodToArray* module was used to crop 256×256 pixel tiled samples following common data science practices (Ronneberger et al., 2015). We collected 20,000 tile samples for training and 5,000 for validation for both *U-Net-A* and *U-Net-B*. Each sampled tiled image included three-layer input data (e.g., *U-Net-A* included urban land-use map for 1994, elevation, and slope) and a single-layer future urban land-use map (e.g., *U-Net-A* included urban land-use map for 2006).

Both U-Net models were trained for 200 epochs (an epoch refers to the U-Net model completely updating its weights using all 20,000 training samples). We saved the model produced at each epoch and tested its performance on the 5,000 validation samples. During the training process, the tiled input images were resized to 8×8 pixels after five down-sampling operations and then converted to a single-layer output image that was the same size as the original input image tile with another five up-sampling operations (Fig. 4). During the validation process, the mean squared error (MSE) was used to compute the difference between the output and target images (i.e., the loss) of the model. Finally, the model with the lowest MSE was determined to be the best model for projecting the target urban layout.

2.6. Producing the simulation map

The input data (e.g., for *U-Net-A* a three-layer raster of urban land-use for 2006 and the elevation/slope data) was split into tiles, supplied to the trained U-Net model to produce separate outputs, and then the outputs were mosaicked into a single transition potential map. The pixel values of the transition potential map ranged from 0 to 1, indicating the probability of being an urban pixel at the desired projection date. To reduce the tile edge effects, image tiles were cropped to a size that was 32-pixel larger than the training image tiles (Fig. 5). Although the U-Net model was trained on image tiles of 256×256 pixels, it could process the buffered 320×320 (original size of 256 plus buffers at edges of size 32×2) image tiles because of the dimensional insensitivity of CNN structures (Long et al., 2015).

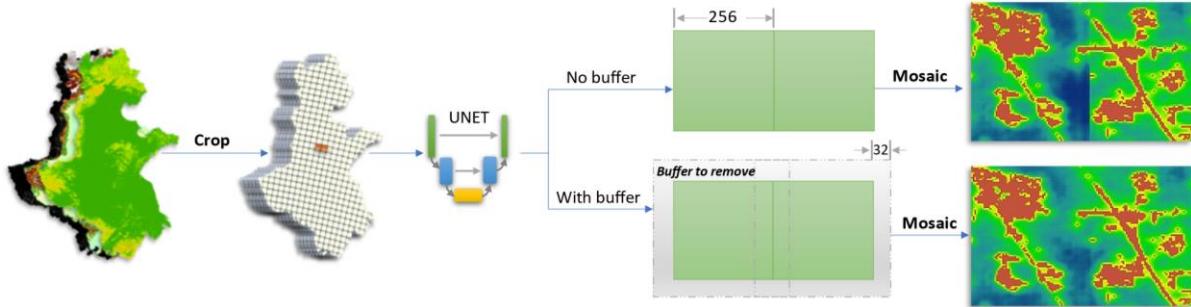


Fig. 5. Buffering the image tile to reduce the edge effects. Because the final output was a mosaic of image tiles, we buffered each image tile with an additional 32 pixels at the edges and removed these buffers before mosaicking to the final output map to reduce the edge effect. A larger buffer size would better alleviate the edge effect but lead to a higher computational cost; thus, we selected a 32-pixel buffer size as it is commonly used in deep learning structures that balance effects and cost (Long et al., 2015).

A classified urban land-use map was created by binarizing the transition potential map using a threshold. First, we ranked the pixel values of the transition potential map from highest to lowest, accumulated the pixel count in this order, and classified them to a value of 1 (indicating urban land-use pixels) until the accumulated count value matched the urban pixel number in the target year. We then allocated the remaining pixels a value of 0 (indicating non-urban land-use).

The historical urban areas were derived using the data from Wang et al. (2021b). Predicted urban areas are listed by province in Table 4 and an exponential extrapolation (Fig. 6) was carried out to binarize the transition potential map for 2030. To allocate the total projected urban land-use area to each province we assumed that the share of urban area within each province remained the same as in 2018, and then binarized each province independently to reduce the bias caused by different regional development levels.

Table 4. Predicted extents of urban areas in each province in 2030.

Province	2018 (km ²)	2030 (km ²)	Increase (%)
Anhui	21070.96	29463.61	39.83
Beijing	2630.09	3338.75	26.94
Hebei	20454.29	26440.38	29.27
Henan	28282.24	38850.77	37.37
Jiangsu	24210.08	33731.72	39.33
Shandong	32005.14	42758.61	33.60
Tianjin	2925.68	3754.65	22.08
Total	131578.48	178338.49	35.54

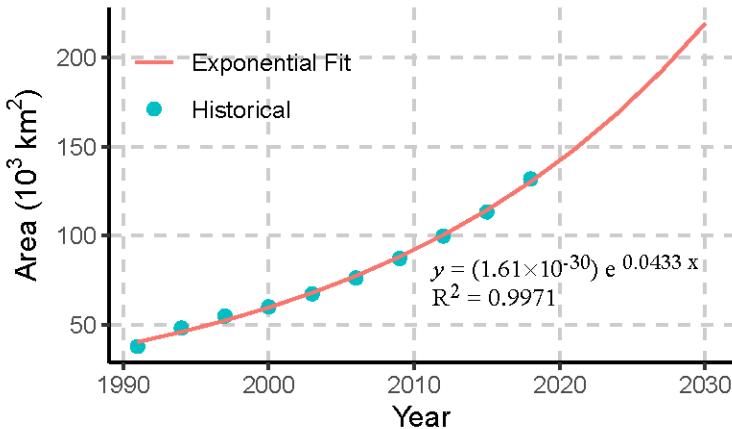


Fig. 6. Exponential regression on historical urban areas in the North China Plain. The historical urban areas are computed using the data from Wang et al. (2021b).

2.7. Validation and accuracy assessment

We selected a series of accuracy metrics and spatial pattern metrics to quantitatively assess the ability of U-Net to accurately project the spatial distributions and patterns of urban areas in the study area. The transition potential map was assessed using the area under the curve (AUC) of the receiver operating characteristic curve (ROC), which illustrates its diagnostic ability to discriminate urban and non-urban land-use pixels under various thresholds (Fawcett, 2006). The predicted urban land-use map was evaluated via map overlay and landscape-level spatial pattern metrics (McGarigal and Marks, 1995). The map-overlay metrics selected in this study were overall accuracy (OA), hit rate, and FoM because they not only quantified how the prediction agrees with the reference but also assessed how the reference was erroneously misrepresented by the prediction. The spatial pattern metrics selected in this study were patch number (PN) and landscape shape index (LSI) because they covered the general landscape features and were capable of assessing the typical shapes and spatial patterns of urban development. The selected metrics are described more fully in Table 5. Validation was performed for each prefecture independently, yielding a total of 76 records for each metric.

Table 5. Validation metrics for evaluating the classified urban land-use maps.

Name	Equation	Explanation
AUC	$\int_{x=0}^1 TPR(FPR^{-1}(x))dx$	AUC measures a model's aggregated performance under different thresholds to discriminate between urban and nonurban land-use pixels (Tong and Feng, 2020).
OA	$(A + D) / (A + B + C + D)$	<i>Overall accuracy</i> is the ratio of correctly identified urban and nonurban land-use pixels to the total number of predictions.
Hit rate	$A / (A + C)$	<i>Hit rate</i> is the ratio of correctly identified urban land-use pixels (i.e., hits) to the number of urban land-use pixels in the reference map.

1	FoM	$A / (A + B + C)$	<i>FoM</i> is the ratio of the intersection to the union upon overlaying predicted urban land-use pixels with reference urban land-use pixels (Pontius et al., 2008).
2	PN	n	<i>Patch number</i> is the number of patches in the urban landscape.
3	LSI	$\frac{\sum p_i}{4\sqrt{\sum a_i}}$	<i>Landscape shape index</i> reflects the complexity of urban landscape patches. For example, a square patch is simple (low value), whereas a linear patch is complex (high value).

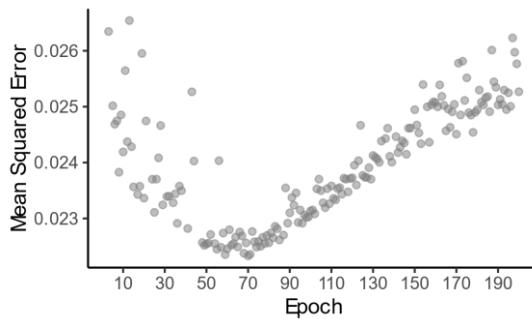
11
 265 Note: *TPR* is the true positive rate, and *FPR* is the false positive rate of the ROC from the transition potential map. A
 266 represents the correctly predicted urban land-use pixels (hit), B represents the incorrectly predicted urban land-use pixels
 267 (false alarm), C represents the incorrectly predicted nonurban land-use pixels (miss), and D represents the correctly
 268 predicted nonurban land-use pixels (correct rejection). n is the total number of landscape patches, p_i is the perimeter
 269 length (m) of the i th path of the urban patches, and a_i is the area (ha) of the i th urban patch.

18 270 Finally, we selected 11 cities of varying sizes from across the study area for visual inspection to
 19 271 qualitatively evaluate the ability of U-Net to simulate realistic spatial urban land-use patterns and
 20 272 development characteristics.

23 273 3. Results

25 274 3.1. Model training

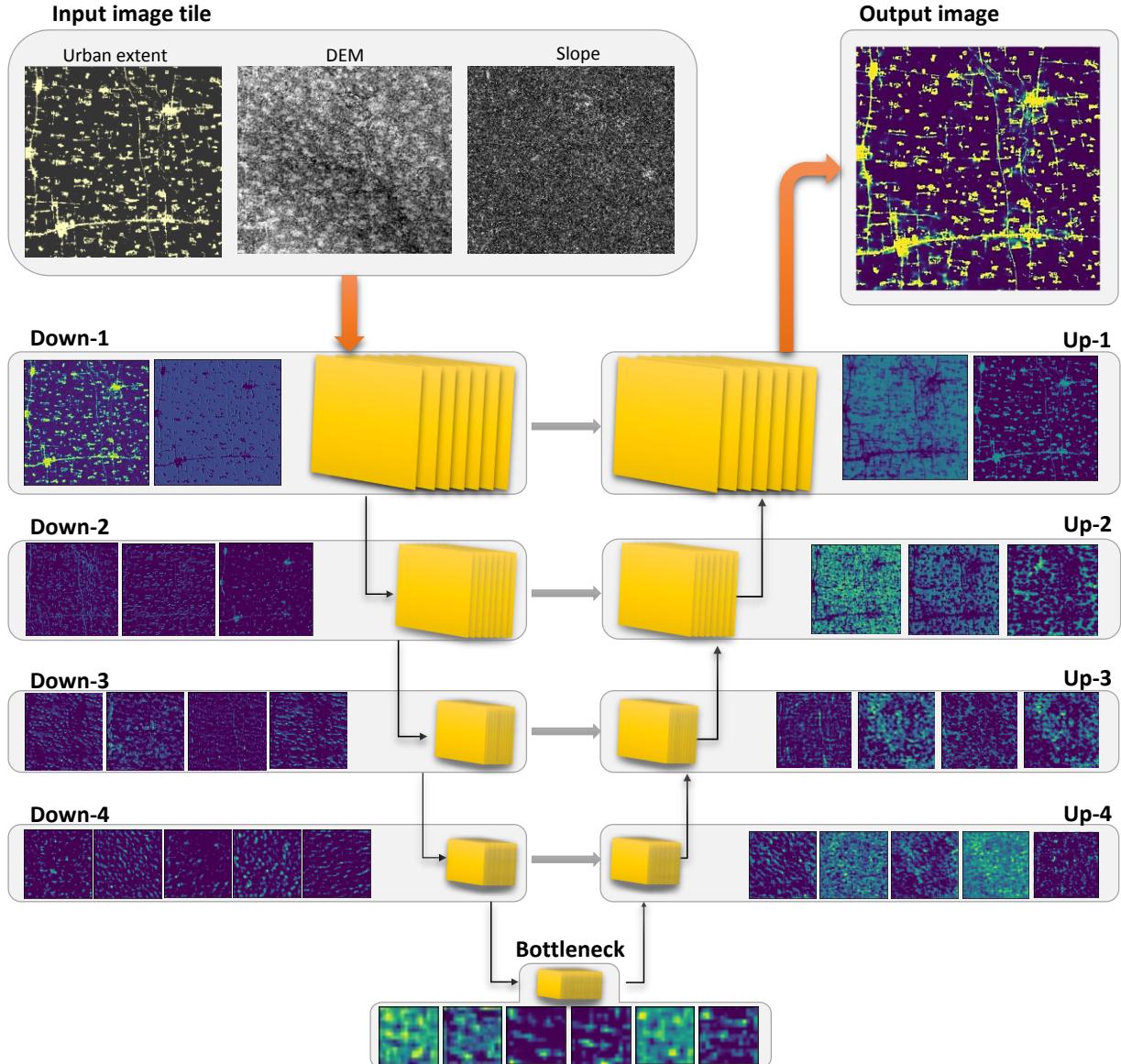
29 275 The MSE of the *U-Net-A* model trained on the urban land-use maps of 1994 and 2006 is shown in Fig.
 30 276 7. The lowest MSE was 0.022 at the 70th epoch, which was determined as the best training epoch to
 31 277 simulate urban land-use for 2018.



43 278 Fig. 7. Mean squared error (MSE) of *U-Net-A* trained on historical urban land-use maps for 1994 and 2006 under different
 44 279 number of training epochs.

47 281 *U-Net-A* successfully learned and captured different aspects of the complex spatial patterns of urban
 48 282 land-use in the study area (Fig. 8). For example, in the illustrative sample presented in Fig. 8, the
 49 283 Down-1 layer broadly distinguished between urban and non-urban land-use pixels. The Down-2 layer
 50 284 recognized simple patterns such as horizontal and vertical roads and learned to distinguish larger
 51 285 towns from small villages. The Down-3 layer learned to associate adjacent urban clusters and
 52 286 identify potential urban development corridors connecting discrete towns/villages. The Down-4
 53 287 layer allocated higher urban development probability to the pixels near existing towns/villages,
 54 288 making urban development corridors more concentrated and intensive. The Bottleneck layers
 55 289 captured the general pattern of urban development in the target year. The up-sampled layers
 56 290 integrated both high-level and low-level features. The Up-4 layer refined the overview patterns in

291 the Bottleneck layer by combining the latter with the Down-4 layer. The Up-3 layer further refined
 292 the spatial features in the Up-4 layer by assimilating urban development corridors identified in the
 293 Down-3 layer. Lastly, the Up-1 layer associated the urban/nonurban feature maps in the Down-1
 294 layer with the Up-2 layer, producing the final output image tile, which allocated more urban land-use
 295 pixels around larger towns while maintaining the refined patterns.



297 Fig. 8. Visualization of an image tile process by different *U-Net-A* layers. The last activation map of each layer was used to
 298 visualize its pattern recognition capability. Note that we selected only a few activation maps for the visualization, given
 299 limited page space.

300 3.2. Validation maps

301 The transition potential map (Fig. 9) illustrates the gravity effect where pixels near larger towns and
 302 cities were associated with higher urban transition potential than those near smaller villages. For
 303 example, a large area surrounding Suixian was identified as having high transition potential while
 304 much smaller areas at the edges of nearby villages exhibited high transition potential. Similar
 305 patterns were found near Yangyuan and Duling. Linear development patterns were also captured

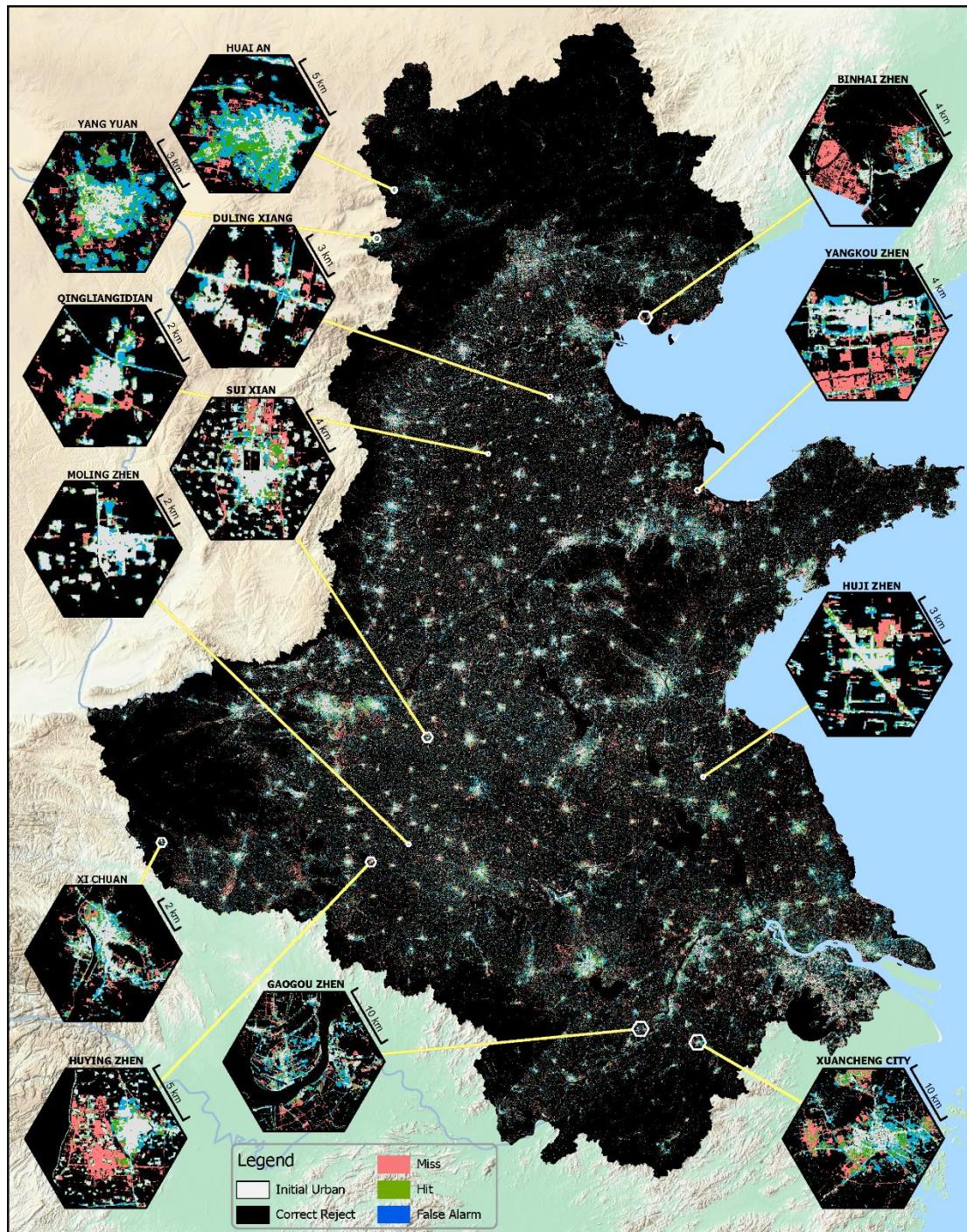
well in the transition potential map. For example, development along roads in Yangkou Zhen and Huiji Zhen were correctly identified despite only appearing as discrete linear segments in the initial urban land-use map for 2006.



Fig. 9. Urban transition potential for 2018 based on the 2006 urban land-use map.

The classified urban land-use map (Fig. 10) created by binarizing the transition potential map correctly identified new urban areas (i.e., hits), particularly those adjacent to existing cities, demonstrating its ability to capture the effect of neighborhood on urban development. It also incorrectly identified some areas as new urban areas (i.e., false alarms) and failed to identify some new urban areas (i.e., misses), particularly in planned new developments. Many of the missed urban areas occurred some distance from the original urban areas, such as the newly developed areas of Binhai Zhen, Yangkou Zhen, and Xuancheng City that were separated from the old town centers.

318 Importantly, the model was exceptionally good at identifying urban expansion along linear features
1 319 such as the new development along major roads in Duling Xiang, Gaogou Zhen, and Huiji Zhen,
2 320 despite the absence of explicit accessibility variables in the input dataset.



321 Fig. 10. Classified urban land-use map including hits, misses, and false alarms.
322

323 3.3. Validation metrics

324 A comparison of accuracy metrics between the output maps and the reference map for 2018 is
325 shown in Fig. 11. The AUC of the transition potential map ranged from 0.70 to 0.97, with a median
326 value of 0.81. The median values of the OA, hit rate, and FoM of the classified urban land-use map
327 were 0.91, 0.33, and 0.20, respectively.

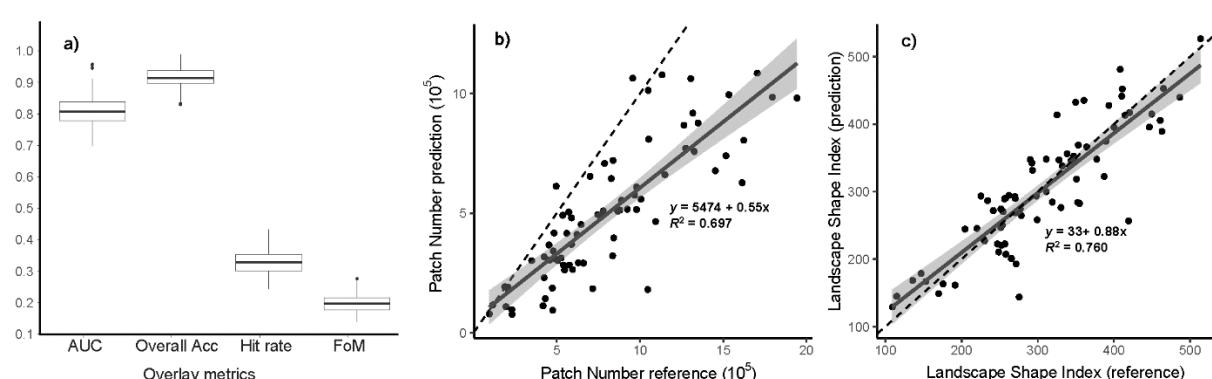


Fig. 11. Assessment metrics for the urban prediction in 2018: a) accuracy metrics; b) patch number; and c) landscape shape metrics. Each dot refers to a record computed from a prefecture in the study area. The dashed line refers to $y = x$ and the ribbon is the confidence interval (95%) of the mean for the fitted linear regression.

The landscape-scale spatial pattern metrics of the classified urban land-use map compared to the reference map are also shown in Fig. 11. The patch number of the prediction map was systematically lower than that of the reference map, while the LSIs of both maps were very similar. This suggests that U-Net tended to predict fewer, more connected urban land patches than the actual urban development depicted in the reference map but produced urban shapes very close to the real-world urban patterns.

3.4. Predicted urban area for 2030

U-Net-B was developed with the same process as *U-Net-A* except for taking urban land-use in 2006 and 2018 for training and predicting urban land-use for 2030 (Fig. 12). Newly predicted urban areas tended to be concentrated around large cities. In Suixian, for example, the predicted urban expansion was much larger around large cities/towns than that around smaller villages. The predicted urban area followed specified patterns rather than sprawling in all directions. For example, the predicted urban land areas in Duling Xiang and Yangkou Zhen filled in gaps in the spatial distribution of existing urban areas and maintained the general shape of the original urban layout. Urban prediction for Fengning County followed the city's elongated development trend, and urban lands emerged along transport routes and at major road intersections, as illustrated by the cases of Huji Zhen and Gaogou Zhen.

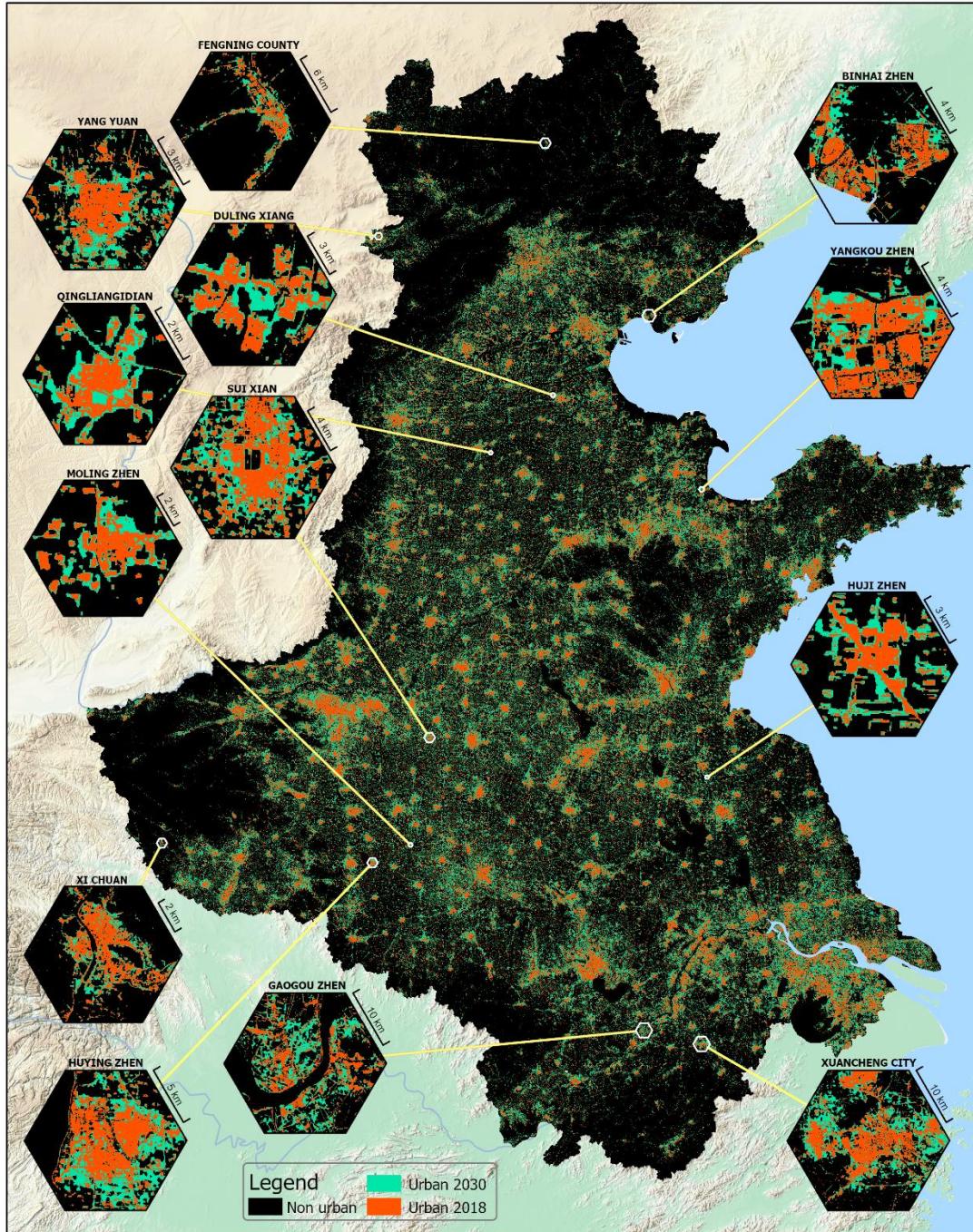


Fig. 12. Predicted urban land areas in 2030.

4. Discussion

4.1. Mimicking real-world urban patterns and dynamics

U-Net was able to model urban land-use change at high accuracy as well as capture and assimilate high-level spatial patterns of urban development in the North China Plain. First, the model captured neighborhood effects. The transition potential map revealed that lands near existing urban areas were more likely to transition to urban land-use in the future and that urbanization amongst larger expanses of non-urban land would be unlikely (Fig. 9). Second, U-Net was able to assimilate small-scale neighborhood effects into large-scale gravity effects. Larger areas of land near larger cities

were predicted to become urbanized in the future, while much smaller areas were predicted to become urbanized around smaller, more remote villages (Fig. 10). Third, U-Net, rather than just simply buffering spatial features, captured the tendency of urban expansion to follow linear patterns. For example, the elongated development trend of Fengning County controlled by valley terrain was identified in the prediction (Fig. 12), and areas along major transport routes were allocated high transition potentials (Fig. 9). However, the U-Net model was unable to predict new planned developments that sprouted up some distance from existing urban areas, such as Binhai Zhen, Yangkou Zhen, Huying Zhen, and Xuancheng City (Fig. 10).

4.2. Reducing subjectivity by automatically constructing transition rules

The large-scale spatial patterns identified by U-Net, in particular the gravity effect of large cities, would not have been captured by existing CA models. A large neighborhood is required for a CA model to incorporate information at a large scale which tends to decrease simulation performance. Wang et al. (2021a) found that the best neighborhood size to simulate urban development in Beijing was 25×25 pixels, while Roodposhti et al. (2020) observed that a 9×9 pixel neighborhood outperformed other settings. However, the unique design of the U-Net model is capable of incorporating more extensive information than previous CA model structures, enabling the neighborhood influence to be learned from the data and urban development to be simulated with more refined spatial configurations.

U-Net did not require extensive amounts of forcing data or parameterization to simulate urban land-use changes, except for the learning rate and the size of the training images. The transition rules were automatically generated in the form of learned weights, and training was an automatic process that required few subjective decisions to be made. In contrast, CA model structures have often required a complex calibration process to determine suitable parameters. For example, Feng and Tong (2020) developed a framework that integrated three algorithms and different neighborhood settings to simulate urban growth in Shanghai, whereas an array of parameters including inertia weights, decay magnitude, spatial heterogeneity, and variable scaling, needed to be specified before running the model. Many studies determine these parameters according to expert knowledge (Mustafa et al., 2017; Chen et al., 2020; Tripathy and Kumar, 2019), potentially leading to subjective biases. Other studies went through a systematic parameter selection process to find the best parameters (Roodposhti et al., 2020; Yu et al., 2021), but this process is time-consuming and impractical given the many unknown parameters.

4.3. Accurate prediction and robustness in capturing spatial patterns

There are challenges to comparing model performance across studies. Pontius et al. (2008) reported two factors that profoundly influence urban land-use simulation accuracy: 1) the area of urban expansion, and; 2) the spatial resolution. A positive relationship exists between the FoM and observed land-use change. Prediction errors vanished when the simulation maps were resampled into coarser resolution (Pontius et al., 2008). Given these sensitivities, to enable a fair comparison of the accuracy of our U-Net model outputs with CA model outputs, we selected two CA-based studies with similar historical urbanization areas and the same 30m spatial resolution as our study. Wang et al. (2021c) used a particle swarm optimization algorithm, iterated a range of parameter settings such as a self-recognition component, a social component, inertia weights, and the number of particles,

401 to finally arrive at the best model with an FoM of 0.193 for Zhuji, China. Wang et al. (2021a)
1 402 developed 17 sub-models incorporating four periods of historical urban land-use and tested eight
2 403 different neighborhood sizes (5*5 to 41*41) for each sub-model to ultimately identify the best
3 404 simulation which achieved an FoM of 0.219 for Beijing, China. The FoM (computed from 76
4 405 prefectures) in our study ranges from 0.177 - 0.215 (interquartile range) which covers the 0.193
5 406 reported by Wang et al. (2021c) but is slightly lower than the 0.219 found by Wang et al. (2021a).
6 407 This higher FoM reported by Wang et al. (2021a) may be explained by the very large amount of
7 408 urbanization in their study area (i.e., Beijing) (Pontius et al., 2008). This comparison demonstrates
8 409 that U-Net achieved similar predictive accuracies to comparable CA-based urban land-use models.
9
10
11
12

13 410 4.4. Predicting urban land-use change for formulating policy

14 411 An additional area of 46,760 km² was predicted to become urbanized in the North China Plain
15 412 between 2018 and 2030. The three provinces with the highest rates of urban area increase (Anhui,
16 413 Jiangsu, and Henan) were all in the southern part of the region (Fig. 2). Our projections for 2030 also
17 414 captured urban development in the three megacity groups (Beijing-Tianjin-Hebei, the Yangtze River
18 415 Delta, and the Central Plains) that account for one-third of China's GDP (National Bureau of Statistics
19 416 of China, 2019b). China's strategic development planning process can benefit from the predictions
20 417 arising out of this study in many ways. For example, in urban area predictions can be used to plan for
21 418 infrastructure to support socio-economic development. Simulations can assist policy formulation
22 419 that is tailored to the expected rate, location, and patterns of urbanization.
23
24
25
26
27

28 420 In addition, the competition between urban development and food production in the study area is a
29 421 critical issue for China's food security (Jin et al., 2019). Pressure on arable land could be alleviated by
30 422 making use of our highly granular predictions of future urban development because they enable the
31 423 development of plans/policies to reduce food production losses that would otherwise arise from
32 424 urban expansion (Jin et al., 2019). Ecosystem protection, biodiversity, and environmental
33 425 conservation can also greatly benefit from an accurate prediction of future urbanization. For
34 426 example, urban land is an important proxy for domestic water consumption (Hoekstra et al., 2018)
35 427 and pollution (Zeller et al., 2019), which can be quantified based on urban land-use maps. Finally,
36 428 future urban layouts provide a baseline for local governments to address land-use conflicts and
37 429 settle competing interests among different stakeholders in the urbanization process.
38
39
40
41
42

43 430 4.5. Limitations and prospects

44 431 Spatial patterns of urban development are highly complex. Very few tools currently exist to
45 432 quantitatively compare these patterns. The shape index employed in this study revealed a high level
46 433 of agreement between the U-Net projection of urban areas and the reference land-use map.
47 434 However, our assessment of the ability of the U-Net to mimic specific patterns was largely based on
48 435 visual inspection. No quantitative metrics currently exist to objectively assess whether projected
49 436 urban patterns and shapes look plausible. Expert human interpretation such as the visual inspection
50 437 method used in this study can effectively assess the realism of projection urban patterns but it is a
51 438 qualitative and imprecise process. Better tools and metrics are required to quantitatively assess
52 439 these complex spatial shapes and patterns to complement qualitative assessment based on human
53 440 interpretation.
54
55
56
57
58
59
60
61
62
63
64
65

1 441 U-Net automatically learned the spatial patterns from historical urban land-use change, while CA
2 442 methods model urban development based on the influence of the driving factors and known
3 443 urbanization processes. As a result, U-Net has a limited capacity to support participatory modeling
4 444 that requires a series of ‘what-if’ trials undertaken with stakeholders involving the tweaking of the
5 445 model parameters and driving factors and inspection of the results. Conversely, the U-Net model
6 446 focuses on predictive accuracy rather than supporting participatory modeling. The transition rules of
7 447 U-Net are hidden in its large number of learned weights and are not human-readable, whereas CA
8 448 models are usually explainable and intuitive, i.e., the key driving factors can be identified. Hence, the
9 449 utility of U-Net as a participatory planning tool, where co-learning of the driving factors of
10 450 urbanization is a key objective, is limited at present. However, emerging developments in extracting
11 451 human-recognizable knowledge from deep learning architectures (Gunning et al., 2019; Dosilovic et
12 452 al.) could be applied to U-Net models for a better understanding of the mechanisms behind urban
13 453 land-use dynamics.
14

15 454 Several new, advanced deep learning structures built on computer vision technology are emerging
16 455 which could be used to retrieve a wide spectrum of features (e.g., distance to roads, and/or the
17 456 temporal characteristics in historical urban development) and improve simulation performance.
18 457 There is significant potential to explore the ability of other deep learning architectures to accurately
19 458 model the spatial distribution and pattern of urban development.
20

21 459 **5. Conclusion**

22 460 For our case study region of the North China Plain, the U-Net deep learning architecture was able to
23 461 accurately simulate urban development as well as reveal key urbanization processes. The U-Net
24 462 model successfully captured neighborhood effects whereby new urban areas are more likely to arise
25 463 near existing urban areas. U-Net also captured gravity effects where new urban development is
26 464 more likely to occur near large cities rather than small villages, and it captured the tendency for
27 465 linear expansion of urban land-use along transportation routes but it was not able to predict new
28 466 planned urban development distant from existing urban areas. Nonetheless, the spatial patterns of
29 467 simulated urban land-use maps matched the reference map closely, indicating that U-Net was
30 468 capable of simulating urban land-use change at a highly granular level. In addition, the model had a
31 469 very low requirement for parameterization thereby limiting the subjective decisions and biases in
32 470 training the model. It also required very little forcing data making it particularly suitable for data-
33 471 scarce regions. The more accurate identification of urban patterns and dynamics is a useful addition
34 472 to urban land-use simulation studies to incorporate the various complex spatial driving factors. The
35 473 resulting projected 2030 urban land-use map provides key information for planning China’s strategic
36 474 socioeconomic development and can benefit policy formulation and decision-making concerning
37 475 food security, biodiversity conservation, and environmental protection.
38

39 476 **References**

- 40 477 Agarap, A.F. (2018). Deep Learning using Rectified Linear Units (ReLU).
41 478 Akın, A., & Erdoğan, M.A. (2020). Analysing temporal and spatial urban sprawl change of Bursa city
42 479 using landscape metrics and remote sensing. *Modeling Earth Systems and Environment*, 6, 1331–
43 480 1343.

- 481 Carneiro, M.G., & Oliveira, G.M.B. (2013). Synchronous cellular automata-based scheduler initialized
1 by heuristic and modeled by a pseudo-linear neighborhood. *Natural Computing*, 12, 339–351.
2
3 Chen, S., Feng, Y., Tong, X., Liu, S., Xie, H., Gao, C., & Lei, Z. (2020). Modeling ESV losses caused by
4 urban expansion using cellular automata and geographically weighted regression. *The Science of
5 the total environment*, 712, 136509.
6
7 Clarke, K.C., & Johnson, J.M. (2020). Calibrating SLEUTH with big data: Projecting California's land use
8 to 2100. *Computers, Environment and Urban Systems*, 83, 101525.
9
10 Dosić, F.K., Brcic, M., & Hlupic, N. Explainable artificial intelligence: A survey, 210–215.
11
12 Fan, H., Zhao, C., & Yang, Y. (2020). A comprehensive analysis of the spatio-temporal variation of
13 urban air pollution in China during 2014–2018. *Atmospheric Environment*, 220, 117066.
14
15 Fan, P., Chen, J., Ouyang, Z., Groisman, P., Loboda, T., Gutman, G., Prishchepov, A.V., Kvashnina, A.,
16 Messina, J., Moore, N., Myint, S.W., & Qi, J. (2018). Urbanization and sustainability under
17 transitional economies: a synthesis for Asian Russia. *Environmental Research Letters*, 13, 95007.
18
19 Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874.
20
21 Feng, Y., & Tong, X. (2020). A new cellular automata framework of urban growth modeling by
22 incorporating statistical and heuristic methods. *International Journal of Geographical
23 Information Science*, 34, 74–97.
24
25 Gantumur, B., Wu, F., Vandansambuu, B., Tsegmid, B., Dalaibaatar, E., & Zhao, Y. (2020).
26 Spatiotemporal dynamics of urban expansion and its simulation using CA-ANN model in
27 Ulaanbaatar, Mongolia. *Geocarto International*, 1–16.
28
29 Gao, C., Feng, Y., Tong, X., Lei, Z., Chen, S., & Zhai, S. (2020). Modeling urban growth using spatially
30 heterogeneous cellular automata models: Comparison of spatial lag, spatial error and GWR.
31
32 *Computers, Environment and Urban Systems*, 81, 101459.
33
34 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth
35 Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202.
36
37 Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI-Explainable artificial
38 intelligence. *Science robotics*, 4.
39
40 Hoekstra, A.Y., Buurman, J., & van Ginkel, K.C.H. (2018). Urban water security: A review.
41
42 *Environmental Research Letters*, 13, 53002.
43
44 Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing
45 Internal Covariate Shift. *International Conference on Machine Learning*.
46
47 <http://proceedings.mlr.press/v37/ioffe15.html>.
48
49 Ji, S., Wei, S., & Lu, M. (2019). A scale robust convolutional neural network for automatic building
50 extraction from aerial and satellite imagery. *International Journal of Remote Sensing*, 40, 3308–
51 3322.
52
53 Jin, G., Chen, K., Wang, P., Guo, B., Dong, Y., & Yang, J. (2019). Trade-offs in land-use competition
54 and sustainable land development in the North China Plain. *Technological Forecasting and Social
55 Change*, 141, 36–46.
56
57 Kafy, A.-A., Naim, M.N.H., Subramanyam, G., Faisal, A.-A., Ahmed, N.U., Rakib, A.A., Kona, M.A., &
58 Sattar, G.S. (2021). Cellular Automata approach in dynamic modelling of land cover changes
59 using RapidEye images in Dhaka, Bangladesh. *Environmental Challenges*, 4, 100084.
60
61 Kapinchev, K., Bradu, A., Barnes, F., & Podoleanu, A. (2015). GPU implementation of cross-
62 correlation for image generation in real time, 1–6.
63
64
65

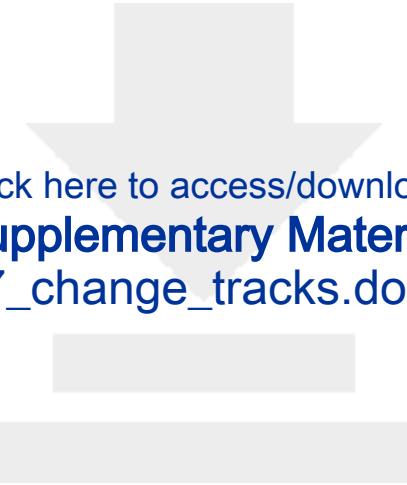
- 525 Kipfer, S. (2018). Pushing the limits of urban research: Urbanization, pipelines and counter-colonial
1 politics. *Environment and Planning D: Society and Space*, 36, 474–493.
2 Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2017). ImageNet classification with deep convolutional
3 neural networks. *Communications of the ACM*, 60, 84–90.
4 Li, X., Gong, P., Le Yu, & Hu, T. (2017). A segment derived patch-based logistic cellular automata for
5 urban growth modeling with heuristic rules. *Computers, Environment and Urban Systems*, 65,
6 140–149.
7 Liu, Y., Batty, M., Wang, S., & Corcoran, J. (2021). Modelling urban change with cellular automata:
8 Contemporary issues and future research directions. *Progress in human geography*, 45, 3–24.
9 Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation,
10 3431–3440.
11 Mansour, S., Al-Belushi, M., & Al-Awadhi, T. (2020). Monitoring land use and land cover changes in
12 the mountainous cities of Oman using GIS and CA-Markov modelling techniques. *Land use policy*,
13 91, 104414.
14 McGarigal, K., & Marks, B.J. (1995). *FRAGSTATS: spatial pattern analysis program for quantifying
15 landscape structure. Spatial Pattern Analysis Program for Quantifying Landscape Structure*.
16 Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.
17 Motlagh, Z.K., Lotfi, A., Pourmanafi, S., Ahmadizadeh, S., & Soffianian, A. (2020). Spatial modeling of
18 land-use change in a rapidly urbanizing landscape in central Iran: integration of remote sensing,
19 CA-Markov, and landscape metrics. *Environmental monitoring and assessment*, 192, 695.
20 Murray, N., & Perronnin, F. (2014). Generalized Max Pooling.
21 Mustafa, A., Cools, M., Saadi, I., & Teller, J. (2017). Coupling agent-based, cellular automata and
22 logistic regression into a hybrid urban expansion model (HUEM). *Land use policy*, 69, 529–540.
23 Mustafa, A., Heppenstall, A., Omrani, H., Saadi, I., Cools, M., & Teller, J. (2018). Modelling built-up
24 expansion and densification with multinomial logistic regression, cellular automata and genetic
25 algorithm. *Computers, Environment and Urban Systems*, 67, 147–156.
26 National Bureau of Statistics of China (2019a). Announcement of the 2019 grain output.
27 http://www.gov.cn/xinwen/2019-12/07/content_5459250.htm.
28 National Bureau of Statistics of China (2019b). *China Statistical Yearbook*. Beijing, China: China
29 Statistics Press.
30 Newland, C.P., Zecchin, A.C., Maier, H.R., Newman, J.P., & van Delden, H. (2018). Empirically derived
31 method and software for semi-automatic calibration of Cellular Automata land-use models.
32 *Environmental Modelling & Software*, 108, 208–239.
33 Nezla, N.A., Mithun Haridas, T.P., & Supriya, M.H. (2021). Semantic Segmentation of Underwater
34 Images using UNet architecture based Deep Convolutional Encoder Decoder Model, 28–33.
35 Peng, K., Jiang, W., Deng, Y., Liu, Y., Wu, Z., & Chen, Z. (2020). Simulating wetland changes under
36 different scenarios based on integrating the random forest and CLUE-S models: A case study of
37 Wuhan Urban Agglomeration. *Ecological Indicators*, 117, 106671.
38 Planillo, A., Kramer-Schadt, S., Buchholz, S., Gras, P., Lippe, M. von der, & Radchuk, V. (2021).
39 Arthropod abundance modulates bird community responses to urbanization. *Diversity and
40 Distributions*, 27, 34–49.
41 Pontius, R.G., Boersma, W., Castella, J.-C., Clarke, K., Nijs, T. de, Dietzel, C., Duan, Z., Fotsing, E.,
42 Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Mohd Sood, A., Pijanowski, B.,
43 Pithadia, S., Sweeney, S., Trung, T.N., Veldkamp, A.T., & Verburg, P.H. (2008). Comparing the

- 569 input, output, and validation maps for several models of land change. *The Annals of Regional
1 570 Science*, 42, 11–37.
- 2 571 Qian, Y., Xing, W., Guan, X., Yang, T., & Wu, H. (2020). Coupling cellular automata with area
3 572 partitioning and spatiotemporal convolution for dynamic land use change simulation. *The
4 573 Science of the total environment*, 722, 137738.
- 5 574 Qiu, B., Li, H., Tang, Z., Chen, C., & Berry, J. (2020). How cropland losses shaped by unbalanced
6 575 urbanization process? *Land use policy*, 96, 104715.
- 7 576 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019).
8 577 Deep learning and process understanding for data-driven Earth system science. *Nature*, 566,
9 578 195–204.
- 10 579 Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image
11 580 Segmentation, 9351, 234–241.
- 12 581 Roodposhti, M.S., Aryal, J., & Bryan, B.A. (2019). A novel algorithm for calculating transition potential
13 582 in cellular automata models of land-use/cover change. *Environmental Modelling & Software*,
14 583 112, 70–81.
- 15 584 Roodposhti, M.S., Hewitt, R.J., & Bryan, B.A. (2020). Towards automatic calibration of
16 585 neighbourhood influence in cellular automata land-use models. *Computers, Environment and
17 586 Urban Systems*, 79, 101416.
- 18 587 Ruiz Hernandez, I.E., & Shi, W. (2018). A Random Forests classification method for urban land-use
19 588 mapping integrating spatial metrics and texture analysis. *International Journal of Remote
20 589 Sensing*, 39, 1175–1198.
- 21 590 Shafizadeh-Moghadam, H., Asghari, A., Tayyebi, A., & Taleai, M. (2017). Coupling machine learning,
22 591 tree-based and statistical models with cellular automata to simulate urban growth. *Computers,
23 592 Environment and Urban Systems*, 64, 297–308.
- 24 593 Shaw, B.J., van Vliet, J., & Verburg, P.H. (2020). The peri-urbanization of Europe: A systematic review
25 594 of a multifaceted process. *Landscape and Urban Planning*, 196, 103733.
- 26 595 Singh, M., Kumar, B., Rao, S., Gill, S.S., Chattopadhyay, R., Nanjundiah, R.S., & Niyogi, D. (2021). Deep
27 596 learning for improved global precipitation in numerical weather prediction systems.
- 28 597 Tong, X., & Feng, Y. (2020). A review of assessment methods for cellular automata models of land-
29 598 use change and urban growth. *International Journal of Geographical Information Science*, 34,
30 599 866–898.
- 31 600 Tripathy, P., & Kumar, A. (2019). Monitoring and modelling spatio-temporal urban growth of Delhi
32 601 using Cellular Automata and geoinformatics. *Cities*, 90, 52–63.
- 33 602 Valencia, V.H., Levin, G., & Hansen, H.S. (2020). Modelling the spatial extent of urban growth using a
34 603 cellular automata-based model: a case study for Quito, Ecuador. *Geografisk Tidsskrift-Danish
35 604 Journal of Geography*, 120, 156–173.
- 36 605 Wang, H., Guo, J., Zhang, B., & Zeng, H. (2021a). Simulating urban land growth by incorporating
37 606 historical information into a cellular automata model. *Landscape and Urban Planning*, 214,
38 607 104168.
- 39 608 Wang, J., Hadjikakou, M., & Bryan, B.A. (2021b). Consistent, accurate, high resolution, long time-
40 609 series mapping of built-up land in the North China Plain. *GIScience & Remote Sensing*, 58, 982–
41 610 998.
- 42 611 Wang, R., Feng, Y., Wei, Y., Tong, X., Zhai, S., Zhou, Y., & Wu, P. (2021c). A comparison of proximity
43 612 and accessibility drivers in simulating dynamic urban growth. *Transactions in GIS*, 25, 923–947.

- 613 Xia, C., & Zhang, B. (2021). Exploring the effects of partitioned transition rules upon urban growth
614 simulation in a megacity region: a comparative study of cellular automata-based models in the
615 Greater Wuhan Area. *GIScience & Remote Sensing*, 1–24.
- 616 Xing, W., Qian, Y., Guan, X., Yang, T., & Wu, H. (2020). A novel cellular automata model integrated
617 with deep learning for dynamic spatio-temporal land use change simulation. *Computers &*
618 *Geosciences*, 137, 104430.
- 619 Yeh, A.G.-O., & Chen, Z. (2020). From cities to super mega city regions in China in a new wave of
620 urbanisation and economic transition: Issues and challenges. *Urban Studies*, 57, 636–654.
- 621 Yu, J., Hagen-Zanker, A., Santitissadeekorn, N., & Hughes, S. (2021). Calibration of cellular automata
622 urban growth models from urban genesis onwards - a novel application of Markov chain Monte
623 Carlo approximate Bayesian computation. *Computers, Environment and Urban Systems*, 90,
624 101689.
- 625 Zeiler, M.D., & Fergus, R. (2013). Visualizing and Understanding Convolutional Networks.
626 <https://arxiv.org/pdf/1311.2901>.
- 627 Zeller, V., Towa, E., Degrez, M., & Achteren, W.M.J. (2019). Urban waste flows and their potential for a
628 circular economy model at city-region level. *Waste management (New York, N.Y.)*, 83, 83–94.
- 629 Zhai, Y., Yao, Y., Guan, Q., Liang, X., Li, X., Pan, Y., Yue, H., Yuan, Z., & Zhou, J. (2020). Simulating
630 urban land use change by integrating a convolutional neural network with vector-based cellular
631 automata. *International Journal of Geographical Information Science*, 34, 1475–1499.
- 632 Zheng, W., Shen, G.Q., Wang, H., Hong, J., & Li, Z. (2017). Decision support for sustainable urban
633 renewal: A multi-scale model. *Land use policy*, 69, 361–371.

Author statement

Jinzhu Wang: Conceptualization, Data curation, Methodology, Visualization, Writing- Original draft preparation. **Michalis Hadjikakoua:** Investigation, Validation, Writing- Reviewing and Editing. **Richard J.Hewitt:** Investigation, Writing- Reviewing and Editing. **Brett A. Bryan:** Supervision, Writing- Original draft preparation, Writing- Reviewing and Editing, Methodology, Visualization.



Click here to access/download
Supplementary Material
07_change_tracks.docx



Click here to access/download
Supporting File
06_Supplementary_20220517.docx