International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

A maps-to-maps approach for simulating urban land expansion based on convolutional long short-term memory neural networks

Zihao Zhou, Yimin Chen, Xiaoping Liu, Xinchang Zhang & Honghui Zhang

To cite this article: Zihao Zhou, Yimin Chen, Xiaoping Liu, Xinchang Zhang & Honghui Zhang (03 Jan 2024): A maps-to-maps approach for simulating urban land expansion based on convolutional long short-term memory neural networks, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2023.2298296

To link to this article: https://doi.org/10.1080/13658816.2023.2298296

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2023.2298296

RESEARCH ARTICLE

A maps-to-maps approach for simulating urban land expansion based on convolutional long short-term memory neural networks

Zihao Zhou[a], Yimin Chen [a], Xiaoping Liu[a], Xinchang Zhang[b] and Honghui Zhang[c]

[a]School of Geography and Planning, Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China; [b]School of Geography and Remote Sensing, Guangzhou University, Guangzhou, China; [c]Guangdong Guodi Planning Science Technology Co., Ltd., Guangzhou, China

ABSTRACT

Cellular automata (CA) have been prevalently used for the simulation of urban land change. However, how to effectively learn the spatial-temporal dynamics of urban development from time-series data remain an important challenge for CA-based models. To address this issue, we propose a new model for the simulation of urban development based on convolutional long short-term memory (ConvLSTM) neural networks. The core of the proposed model is a sequence of vanilla ConvLSTM cells integrated with the modules of channel attention and contextual embedding. Compared with conventional CA-based models, the proposed ConvLSTM model is more advanced in that it can better leverage the open access annual urban land maps to capture simultaneously the spatial structure and the temporal dependency of historical urban development, and further predict multiple maps of annual development for subsequent years (i.e., Maps-to-Maps). The performance of the ConvLSTM model is evaluated through the case studies in China's three mega-urban regions, and ConvLSTM outperforms other state-of-the-art deep learning architectures at both the pixel level and the coarser grid level. The results also suggest the satisfactory transferability of ConvLSTM in that the model trained in one mega-urban region can be successfully re-used in others without fine tuning.

ARTICLE HISTORY
Received 10 August 2023
Accepted 17 December 2023

KEYWORDS
Urban land expansion; cellular automata; convolutional long short-term memory

1. Introduction

Urban land expansion is the direct outcome of urbanization. Globally, population growth was found the major driver of urban land expansion during 1970-2014, while the importance of Gross Domestic Product growth became greater after 2000 (Mahtta *et al.* 2022). Due to its far-reaching impacts, such as arable land losses (Sumbo *et al.* 2023), biodiversity declines (Simkin *et al.* 2022), and urban heat island effects

CONTACT Yimin CHEN chenym49@mail.sysu.edu.cn
Supplemental data for this article can be accessed online at https://doi.org/10.1080/13658816.2023.2298296.

(Chatterjee and Majumdar 2022), researchers have been devoting their efforts to the modeling of urban land change. Many studies focus on mapping urban land change through remotely sensed imagery and advanced detection methods (Aslam *et al.* 2023). In particular, the power of cloud computing has been exploited along with the increasing volume of satellite images to generate urban land maps that are globally covered, high-resolution, and annually available over a long span of time (Liu *et al.* 2020, Huang *et al.* 2022, Zhou *et al.* 2018, Gong *et al.* 2020). These data products are fundamental to the development of urban land change models, which have been considered helpful for understanding the associated impacts of urban land expansion at various regions and scales (Chen *et al.* 2020, Tong and Feng 2020, Gao and O'Neill 2020, Zhang *et al.* 2023b).

Many of the contemporary urban land change models are developed based on cellular automata (CA) because of their ability to simulate spatial phenomena with simple mechanisms (D'Acci and Batty 2019). When CA are applied to the simulation of urban land change, the major components of a CA-based model usually include a module to evaluate development suitability considering a set of spatial variables (or driving factors), a neighborhood component to model the local interaction between neighboring land units, and a stochastic component to account for the effects of unpredictable perturbation (Roodposhti *et al.* 2020). How these components operate is controlled by the mechanisms represented in the form of either if-then rules or functions that can be calibrated using spatial data. A variety of methods have been proposed for model calibration. They range from geostatistical methods (Gao *et al.* 2020), heuristic methods (Cao *et al.* 2019, Modiri *et al.* 2023), tree-based algorithms (Shafizadeh-Moghadam *et al.* 2021), to the more recent deep learning methods such as convolutional neural networks (CNN) (Zhai *et al.* 2020, Wang *et al.* 2022, Guan *et al.* 2023).

Despite the different calibration methods applied, most of the contemporary CA-based models are calibrated using a pair of urban land maps at two points in time: the initial T_0 and the ending T_1 at which some change has occurred, and validated using the urban land map at T_2 (e.g. Addae and Dragićević (2022)). A few models are calibrated using only the T_0 urban land map, such as FLUS (Liu *et al.* 2017). Therefore, the calibration methods used in these models merely captures the spatial dynamics underlying the process of urban land expansion, but fail to interpret the (spatial-)temporal information from multi-temporal or time-series urban land maps. This is not trivial because ignoring the temporal dimension would cause misunderstanding of the process of urban land expansion and reduce the reliability of the long-run projections of urban land change (Li *et al.* 2021).

To address this issue, several methods have been developed to incorporate multi-temporal or time-series urban land data for model calibration. For instance, Chen *et al.* (2016) used survival analysis to capture the temporal trend of the spatial variables' effects by incorporating four urban land maps with an interval of five years. The major weakness of their approach is that only the constant trend over time (i.e., increase or decrease) can be obtained, neglecting the potential temporal fluctuations. Li *et al.* (2020) proposed to adjust the neighborhood component of a Logistic-CA model by assigning higher weights for the neighboring cells that were developed closer to present time. However, the weights were determined mainly based on expert knowledge and difficult to calibrate with empirical data. Yang *et al.* (2023) focused on delineating the spatial-temporal trend of urban land demand with a isotropic Gaussian function, but ignored the spatial-temporal information associated with the neighbourhood component and the spatial variables, which is inherent in time-series urban land data. Overall, the representative studies mentioned above cannot fully address the challenge of capturing the spatial-temporal dynamics of urban land expansion.

More recently, deep learning methods have emerged as a new approach for modeling urban land expansion with multi-temporal or time-series data. For instance, Qian *et al.* (2020) developed a CNN-based method to learn the effects of spatial variables on urban land expansion from multi-temporal urban land maps through convolution operations. Geng *et al.* (2022) adopted the similar approach but only three urban land maps were used with an interval of five years. Despite the use of multi-temporal data in Qian *et al.* (2020) and Geng *et al.* (2022), they were only stacked as the input while the temporal relationships of the data could not be effectively captured, because the architecture of CNN lacks a component to do that. Another approach of deep learning, namely long short-term memory (LSTM) (Graves 2012), had demonstrated promising performance in the tasks of sequence prediction, hence it is applicable to the spatial-temporal modeling of long-run urban land change. Despite the success of LSTM in many fields, such as natural language processing (Merity *et al.* 2018) and financial market forecasting (Moghar and Hamiche 2020), the development of LSTM-based models for simulating urban land change are still at its infancy. Liu *et al.* (2021) applied the LSTM method for the simulation of urban expansion, but the power of LSTM could not be fully exploited because only three urban land maps (with an interval of ten years) were used. Xing *et al.* (2020) combined CNN and LSTM (i.e., CNN-LSTM) to develop a CA-based simulation model of urban development, in which the CNN component is for learning the effects of spatial variables and the LSTM component is for learning the temporal relationships from the input time-series urban land maps. A similar CNN-LSTM-CA model was also developed by Zhou *et al.* (2023).

This study proposes a new model for the simulation of urban land expansion based on the method of convolutional long short-term memory (ConvLSTM) (Shi *et al.* 2015). ConvLSTM extends the conventional LSTM methods by including the convolutional operations for encoding spatial-temporal information and making prediction. Therefore, compared with the conventional CNN-LSTM method that capture the spatial information and the temporal information using two separating networks (i.e., CNN and LSTM, respectively), ConvLSTM has a more concise architecture that can learn and fuse the spatial-temporal information in the same network (Masrur and Yu 2023). Moreover, unlike the conventional LSTM methods that only yield the prediction at one step ahead, ConvLSTM allows for outputting a sequence of future results (i.e., multiple timesteps). All these characteristics make ConvLSTM more suitable than the conventional methods to the challenging simulation of long-run urban land change.

In this study, we further enhance ConvLSTM with the modules of channel attention (Woo *et al.* 2018) and contextual embedding (Chai *et al.* 2022). Because of its conciseness and efficiency, channel attention has been widely used for weighting the importance of features learned by deep learning models (Cai and Chen 2021). Therefore, we use channel attention to enhance the ability of the original ConvLSTM to estimate the effects of spatial variables. Contextual embedding, however, is originally designed to improve the contextual interactions when applying LSTM to solve the problems of spatial-temporal prediction (Chai *et al.* 2022), and hence is feasible to combine with ConvLSTM for simulating urban land expansion. By this means, the simulation model uses a sequence of urban land maps as the input and yields a sequence of predicted urban land maps as the output (i.e., a Maps-to-Maps manner). The proposed ConvLSTM model was evaluated and compared with several other models through the simulations of urban land expansion in China's Guangdong-Hong Kong-Macau Great Bay Area (GBA). The transferability of the ConvLSTM model was also examined through additional experiments of simulations for Yangtze River Delta (YRD) and Beijing-Tianjin-Tangshan (BTT).

2. Materials and methods

2.1. Study areas and data

The mega-urban regions of GBA, YRD, and BTT have the highest economic and political status in China, and are considered in the Fourteenth Five-Year Plan as the key regions for China's long-run development. During the past three decades, however, the rapid expansion of urban land in these three mega-urban regions has caused the increasing pressure on natural resources and environments. Therefore, accurate simulation of urban development in these regions is not only useful to understand the historical urban dynamics but also important to explore the consequences of future development. In this study, the performance of the proposed model is first evaluated using the GBA data, and the geographic transferability of the proposed model is further assessed using the collective data of all three mega-urban regions.

The urban land maps of the study area were obtained from the 30-m raster dataset of global impervious surface area (GISA 2.0) (Huang *et al.* 2022). This dataset was developed based on the global Landsat images spanning from 1985 to 2019, and provides annual urban land cover with higher accuracy than other 30-m datasets. In this study, the annual urban land maps of the study areas during the periods of 2000-2017 were used for the GBA simulation and the evaluation of transferability. Specifically, the 2000-2005 maps were used to trained the proposed model to produce the simulated 2006-2011 maps, and the independent 2012-2017 maps were used to evaluate the performance of the trained model in terms of prediction.

Besides the urban land maps, the data of DEM, road networks, urban facility, and the positions of representative urban centers were also used. The DEM data was derived from the recently released Forest and Buildings removed Copernicus DEM (FABDEM) with a spatial resolution of 30 m (Hawker *et al.* 2022). The road networks data was collected from OpenStreetMap, which provides the major roads of the study areas. The urban facility considered include airports, railway stations, and accesses to highways. The representative urban centers include county centers and town centers. Data of urban facility and representative urban centers were acquired by digitalizing the official maps of the study area. With these data, several spatial variables were generated, including slope, distance to the nearest major road, distance to the nearest airport, distance to the nearest railway station, distance to the nearest access to highways, distance to the nearest county center, and distance to the nearest town center. All these spatial variables represent the situation of year 2020 and have a 30-m resolution. The selection of these variables is aligned with those in previous studies (Chen *et al.* 2019b, Rienow *et al.* 2021, Roodposhti *et al.* 2020). Note that the spatial variables are held constant throughout the simulation of annual urban growth. Such a treatment is indeed a compromise due to the lack of multi-temporal urban facility and road network data. Recent studies also applied constant spatial variables for urban simulation using multi-temporal or time-series urban land maps (Wang *et al.* 2022, Zhou *et al.* 2023, Qian *et al.* 2020).

2.2. Methods

2.2.1. Overview of the model structure

The proposed model simulates urban development through a Maps-to-Maps manner. This means that the model receives multiple, annually consecutive urban land maps for years $T_1, \ldots, T_n$ as the inputs and outputs a sequence of annually consecutive urban land maps for the future years $T_{n+1}, \ldots, T_{n+m}$. To fulfil this approach, a ConvLSTM neural network integrated with the modules of channel attention and contextual embedding is developed using a set of spatial variables (Table 1). Specifically, the network is composed of a sequence of ConvLSTM cells, each of which is a vanilla ConvLSTM cell (Shi *et al.* 2015) (see Section 2.2.4) enhanced by channel attention (Woo *et al.* 2018) and contextual embedding (Chai *et al.* 2022) (see Section 2.2.2 and 2.2.3, respectively) (Figure 2). The number of the ConvLSTM cells that compose the model is $n + m$ - 1, where n refers to the number of input years and m is the number of output years.

For the first n ConvLSTM cells, they are used for learning the spatial-temporal dynamics of urban land expansion for the years $T_1, \ldots, T_n$. Specifically, the first ConvLSTM cell receives the data tile X_1 as its input to generate the cell state c_1 and the hidden state h_1. Here X_1 is the spatial subset (64×64 or 1.92 km $\times$ 1.92 km) of the urban land map and the spatial variables, while c_1 and h_1 refer to the long-term and the short-term memories, respectively (Yu *et al.* 2019) (see Section 2.2.4). The resulting c_1 and h_1 along with the data tile X_2 are used as the input of the second ConvLSTM cell to generate c_2 and h_2, and so forth until the nth ConvLSTM cell. The resulting h_n of the nth ConvLSTM cell is regarded as the probability of urban land development (referred to as probability tile for short) at the year T_{n+1}, which is further used (instead of the urban land map at the year T_{n+1}) with the spatial variables to compose the data tile X_{n+1} for the next ConvLSTM cell ($n + 1$). The ($n + 1$)th ConvLSTM cell runs in

Table 1. Spatial variables used for simulating urban land expansion.

Spatial variable	Abbreviations	Source data
Slope	Slope	FABDEM
Distance to the nearest road	D2NR	OpenStreetMap
Distance to the nearest airport	D2A	Official maps
Distance to the nearest railway station	D2RS	
Distance to the nearest access to highway	D2H	
Distance to the nearest county center	D2C	
Distance to the nearest town center	D2T	

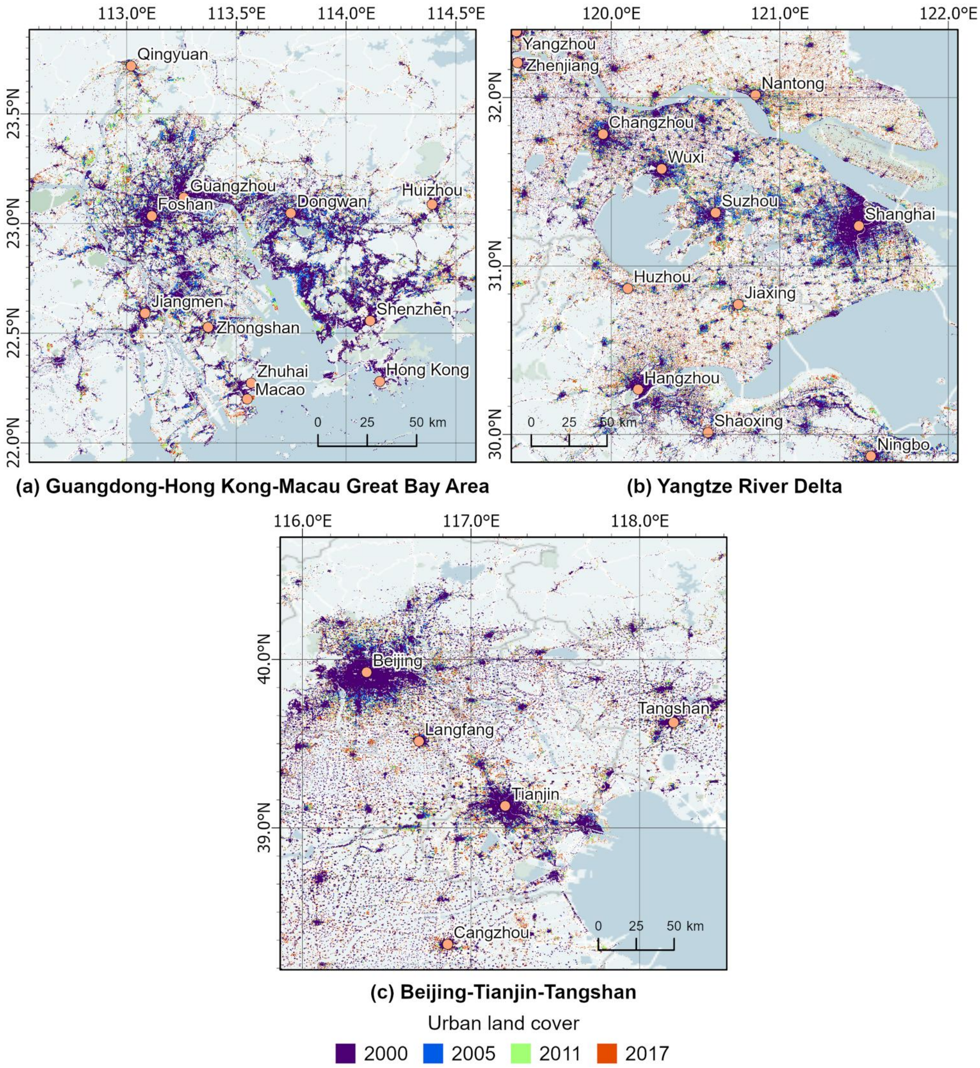

Figure 1. Urban land expansion in the Guangdong-Hong Kong-Macau Great Bay Area (GBA), Yangtze River Delta (YRD), and Beijing-Tianjin-Tangshan (BTT) from 2000 to 2017.

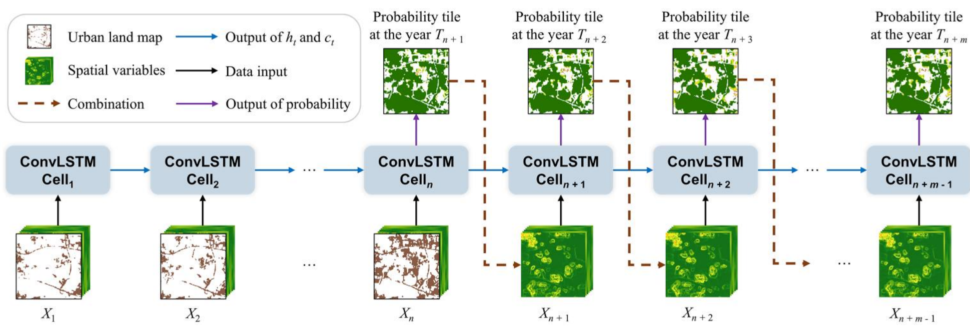

Figure 2. The structure of the ConvLSTM neural network for predicting urban land expansion.

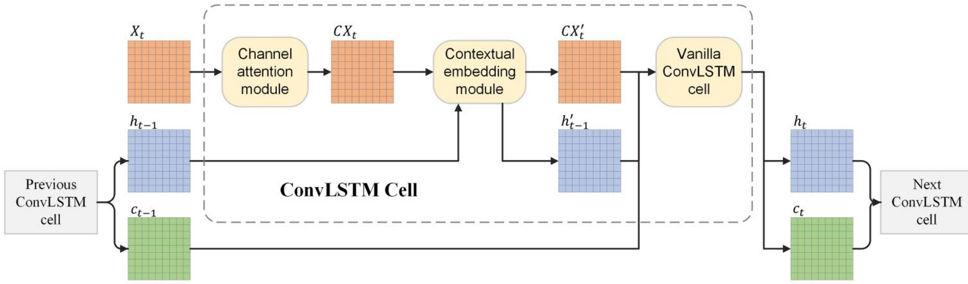

Figure 3. The structure of the proposed ConvLSTM cell.

the same way to generate c_{n+1} and h_{n+1} (as the probability tile at the year T_{n+2}), and so on until the $(n+m$ - 1)th cell, which outputs the probability tile at the year T_{n+m}. Finally, with the resulting probability tiles, urban development for the years $T_{n+1}, \ldots, T_{n+m}$ can be simulated by the means of, for example, value-based segmentation (see Section 2.2.5).

Figure 3 demonstrates the details of a ConvLSTM cell. As aforementioned, a ConvLSTM cell is created based on the integration of the vanilla ConvLSTM cell and the modules of channel attention and contextual embedding. Here the channel attention module is used to capture the effects of each spatial variable, while the contextual embedding model is used to infer where urban development would be most likely to occur. The processing within a ConvLSTM cell can be summarized as follows:

$$CX_t = ChA(X_t) \tag{1}$$

$$CX_t', h_{t-1}' = CE(CX_t,\ h_{t-1}) \tag{2}$$

$$h_t, c_t = VConvLSTM\left(CX_t', h_{t-1}', c_{t-1}\right) \tag{3}$$

where *ChA*(), *CE*(), and *VConvLSTM*() correspond to the operations of the channel attention module, the contextual embedding module, and the vanilla ConvLSTM cell. Specifically, the data tile X_t at the year T_t passes through the channel attention module, generating the output CX_t. The resulting CX_t along with $h_{t\text{-}1}$ become the input of the contextual embedding, and the corresponding outputs along with $c_{t\text{-}1}$ are used to generate c_t and h_t based on the operation of the vanilla ConvLSTM cell. The details of the calculations are discussed in the following sections 2.2.2-2.2.4.

With the annual urban land maps and the spatial variables, the ConvLSTM model is first trained to simulate the annual urban development for the period of 2006-2011 with the actual 2000-2005 data as the input. After that the trained model is further evaluated through its prediction of the annual urban development for the period of 2012-2017 with the actual 2006-2011 data as the input. Therefore, the number of ConvLSTM cells is 11 (i.e., $n = m = 6$). Note that the values of six for both n and m is set considering the balance between the amount of spatial-temporal data used by the model and the computation costs. When n is set with greater values, more years of data can be utilized by the model and hence richer spatial-temporal information acquired, but in turn the computation costs would increase. Similarly, when m is set with greater values, more computation time is required to produce the simulations. Also, there is no need to always set n and m equal.

2.2.2. Channel attention module

The channel attention module consists of three components: the global max pooling operation, the global average pooling operation, and the shared weight multi-layer perceptron (MLP) (Woo *et al.* 2018). Here a channel refers to a specific component of the input data tile (e.g. the urban land map or a certain spatial variable). The purpose of using the channel attention module is to capture the importance of each component in the input data tile X_t and output its weighted form CX_t. Figure 4 depicts the workflow of the channel attention module. The input data tile X_t, which is also called the input feature map, is first transformed separately by the operations of global max pooling and global average pooling:

$$mp_i = \max(channel_i) \tag{4}$$
$$ap_i = \text{average}(channel_i) \tag{5}$$
$$V_{\max} = [mp_1, mp_2, \ldots, mp_{k+1}] \tag{6}$$
$$V_{avg} = [ap_1, ap_2, \ldots, ap_{k+1}] \tag{7}$$

where $channel_i$ refers to the ith channel of X_t, and mp_i and ap_i are the respective results of the global max pooling and the global average pooling for $channel_i$. By this means, the input data tile X_t with $(1+k)$ channels (including the urban land map and k spatial variables) is transformed to two vectors (V_{max} and V_{avg}) of size $1 \times 1 \times (1+k)$.

After the operations of global max pooling and global average pooling, the resulting vectors V_{max} and V_{avg} are used as the inputs to train a shared weight MLP to predict urban development at the year $t+1$. This is for obtaining the weights that represent the importance of each channel of X_t. The shared weight MLP consists of three fully connected layers l_1, l_2, and l_3 with the corresponding sizes of $1 \times 1 \times (1+k)$, $1 \times 1 \times 1$, and $1 \times 1 \times (1+k)$. The parameters to optimize include the weight vector W and the bias vector b. The back-propagation method to update the two parameters is based on the gradients derived from the difference between the model outputs and the observed urban land maps. It should be noted that the back propagation only starts after the forward propagation stages are finished in all ConvLSTM cells' MLPs. More details of such a training approach can be found in Woo *et al.* (2018). After that the weights for V_{max} and V_{avg} are calculated as follows:

$$W_{\max} = \sigma(W_3\sigma(W_2\sigma(W_1V_{\max} + b_1) + b_2) + b_3) \tag{8}$$
$$W_{avg} = \sigma(W_3\sigma(W_2\sigma(W_1V_{avg} + b_1) + b_2) + b_3) \tag{9}$$

where W_1, W_2, and W_3 are the optimized weights in l_1, l_2, and l_3, and b_1, b_2, and b_3 are the optimized bias in l_1, l_2, and l_3; and $\sigma()$ is a sigmoid activation function. The

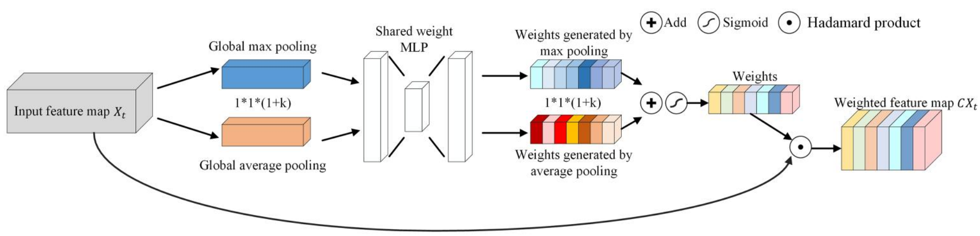

Figure 4. The workflow of the channel attention module.

resulting W_{max} and W_{avg} are then combined as a single weight vector W_X through the following calculation:

$$W_X = \sigma(W_{\max} + W_{avg}) \tag{10}$$

Finally, the input data tile X_t is then weighted through the calculation of Hadamard product:

$$CX_t = weight \odot X_t \tag{11}$$

2.2.3. Contextual embedding module

This module (Chai *et al.* 2022) is used to obtain the contextual information that favors the prediction of urban land expansion using the weighted data tile CX_t and h_{t-1}. The structure of the contextual embedding module is shown in Figure 5(a). The major components of this module are the two convolutional layers that process the input CX_t and h_{t-1}. By this means, the module can capture the interaction between CX_t and h_{t-1}, which is useful to enhance the model's performance in sequence prediction (Melis *et al.* 2019).

Specifically, the input h_{t-1} passes through the first convolutional layer $Conv_h$ and is further transformed by using a sigmoid activation function, yielding the output $h_{t-1,conv}$:

$$h_{t-1,conv} = \sigma(W_{h,conv} * h_{t-1} + b_{h,conv}) \tag{12}$$

where $W_{h,conv}$ and $b_{h,conv}$ are the convolution kernel and bias, respectively. The resulting $h_{t-1,conv}$ is then combined with the input CX_t through the calculation of Hadamard product, and the results are further used as the input of the second convolutional layer $Conv_X$ to generate h'_{t-1}:

$$CX'_t = h_{t-1,conv} \odot CX_t \tag{13}$$

$$h'_{t-1} = \sigma\left(W_{X,conv} * CX'_t + b_{X,conv}\right) \odot h_{t-1} \tag{14}$$

where $W_{X,conv}$ and $b_{X,conv}$ refer to the kernel and bias, respectively, for $Conv_X$. The training processes to update $W_{h,conv}$, $W_{X,conv}$, $b_{h,conv}$, and $b_{X,conv}$ are similar to that of the shared weight MLP. See Chai et al (2022) for more details.

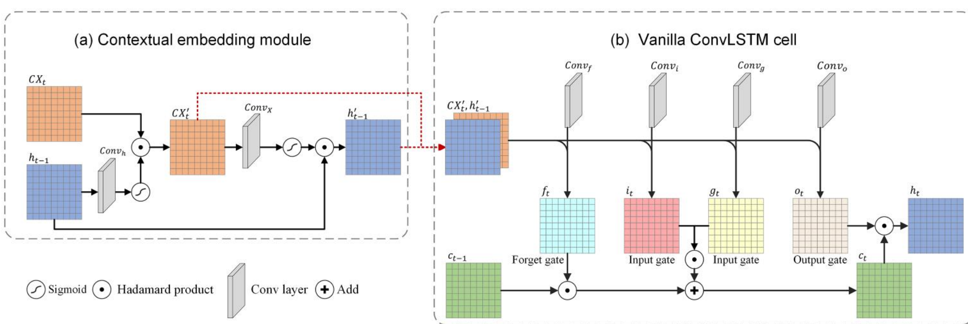

Figure 5. The structures of (a) the contextual embedding module and (b) the vanilla ConvLSTM cell.

2.2.4. Vanilla ConvLSTM cell

The vanilla ConvLSTM cell receives the outputs of the contextual embedding module, based on which the cell state c_t and the hidden state h_t are generated for the current cell (Figure 5(b)). The vanilla ConvLSTM cell consists of four gates that work sequentially. They include a forget gate, two input gates, and an output gate (Yu *et al.* 2019). All these gates are developed using convolutional layers. Specifically, the forget gate determines how much information obtained from the previous ConvLSTM cell should be discarded:

$$f_t = \sigma(W_f * [h'_{t-1}, CX'_t] + b_f) \tag{15}$$

where f_t represents the proportion of information to discard, W_f and b_f are the convolutional kernel and the bias for the convolutional layer $Conv_f$. With the information of the previous ConvLSTM cell being adjusted by f_t, the cell state of the current cell c_{t-1} is then updated through:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{16}$$

$$i_t = \sigma\left(W_i * [h'_{t-1}, CX'_t] + b_i\right) \tag{17}$$

$$g_t = \tanh\left(W_g * [h'_{t-1}, CX'_t] + b_g\right) \tag{18}$$

where i_t and g_t are the outputs of the two input gates; W_i and b_i are the convolutional kernels and bias for $Conv_i$, and W_g and b_g are the convolutional kernels and bias for $Conv_g$. Here the two input gates are used to generate the information that is used to replace the discarded one with the input CX'_t and h'_{t-1}.

Finally, an output gate is implemented to produce the hidden state of the current cell h_t using the input and the resulting c_t:

$$o_t = \sigma\left(W_o * [h'_{t-1}, CX'_t] + b_o\right) \tag{19}$$

$$h_t = o_t \odot \tanh(c_t) \tag{20}$$

where o_t is the result of the output gate; W_o and b_o are the convolutional kernels and bias for $Conv_o$.

2.2.5. Simulation of urban development

The direct outputs of the ConvLSTM model are the annual probability tiles for the years $T_{n+1}, \ldots, T_{n+m}$. Before the simulation of urban development, for each year T, the resulting probability tiles are merged together to form the complete spatial layer of urban development probability. Then the value-based segmentation approach (Feng and Tong 2020, Guan *et al.* 2023) is used to determine the pixels that are developed at the year T. Specifically, for the year T, the non-urban pixels prohibited to develop (e.g., ecologically sensitive areas) are removed. In our case, open water and land units with slope values greater than 25 degrees are considered as ecologically sensitive areas that are not allowed for development. The probability values of the remaining non-urban pixels are then adjusted by random perturbation:

$$P_{i,adj} = P_i \times (1 + (-\ln(\gamma))) \tag{21}$$

where γ is a randomly generated value ranging from 0 to 1. The incorporation of the random perturbantion is to represent the unpredictable impacts of certain stochastic factors on urban development (Roodposhti *et al.* 2020).

After that the set of non-urban cells at the year T are ranked in descending order according to $P_{i,\ adj}$, denoting as $N_{T,\ P}$. With the number of new urban pixels S_T being determined for the year T, the first S_T pixels are selected from $N_{T,\ P}$ as the newly developed urban pixels. An identical simulation process is applied for the year $T+1$ and so forth for the remaining years. For the simulation of historical urban land expansion, the values of S_T for each year are set according to the actual annual urban land maps. It should be noted that besides the value-based segmentation approach, others such as the patch-based simulation strategy (Chen 2022) are also feasible to use based on the resulting probability layers. Nevertheless, we only applied the value-based segmentation approach because testing different simulation methods is not the focus of this study.

The simulations can be validated based on the frequently used indicator of 'Figure-of-Merit' (FoM) (Pontius Jr 2018), which measures the performance of the model in accurately predicting the position of change. Higher values of FoM suggest the better performance of the model under examined. The calculation of FoM is based on the pixel-by-pixel map comparison between the referenced change and the simulated change from T_0 to T_1.

Furthermore, the agreements between the referenced change and the simulated change can also be compared at coarser scales to tolerate the model's prediction error at the pixel level. The comparison is made by aggregating the amounts of the referenced change and the simulated change to the units that are larger than individual pixels, such as moving windows. In this study, we used the moving windows of 1.5 km $\times$ 1.5 km, 3 km $\times$ 3 km, and 6 km $\times$ 6 km to produce the focal sum of newly developed urban land for the references and the simulations, respectively. Regression analysis was then run to reveal how much the simulated change can explain the referenced change. Such a moving-window analysis is a variant of the prevalently used map comparison methods (Hagen-Zanker 2006, Meentemeyer *et al.* 2013, Pijanowski *et al.* 2014) that complement the pixel-by-pixel validation methods with the information of agreement between the simulated and the referenced local composition of land changes.

3. Implementation and results

The proposed model is tested through the simulation of historical urban development in the GBA during 2012-2017. Furthermore, the proposed model is compared against three additional models, namely FLUS, LSTM-CA, and CNN-LSTM-CA. FLUS (Liu *et al.* 2017) features the use of artificial neural networks to estimate development probability, but lacking the ability to handle time series urban land maps. LSTM-CA and CNN-LSTM-CA are developed more recently and represent the current trend of using deep learning techniques to enhance urban CA models with time-series data.

Specifically, LSTM-CA is developed based on the algorithm of LSTM (Staudemeyer and Morris 2019) and focuses on detecting the temporal trend of urban development at the pixel level, considering the sequential states (i.e., the land use type) along the timeline and the effects of spatial variables. CNN-LSTM-CA is developed following the architecture proposed by Xing *et al.* (2020). Indeed, CNN-LSTM-CA is composed of two neural networks, i.e., a CNN and a LSTM network, or can be considered as a LSTM-CA model enhanced by adding a CNN. Here CNN is for estimating the effects of spatial variables at the neighborhood level with the neighborhood size equal to that of the kernels used for convolution. By this means, the spatial information learned by CNN is further combined with the temporal trend captured by LSTM to generate the development probability. Compared with CNN-LSTM-CA, the proposed ConvLSTM model is different in that the convolution operations are embedded directly in the ConvLSTM cells rather than implemented in a separating network as CNN-LSTM-CA does. In that sense, ConvLSTM can simultaneously learn spatial-temporal information from data in the same network. Given such differences, which architecture performs better in the simulation of urban land expansion can be identified through experiments (see Section 3.2).

Section 3.1 explains the training of the ConvLSTM model, while Section 3.2 discusses the differences of performance among different models. Finally, the transferability of the ConvLSTM model is assessed through the simulations of urban development in another two mega-urban regions, namely YRD and BTT, besides the GBA (Section 3.3).

3.1. Settings of model training

The ConvLSTM model was first trained to simulate the urban development during 2006-2011 by using the information of 2000-2005. A total of 6000 data tiles were collected, with 5000 tiles as the training set and the remaining 1000 tiles as the testing set. The initial learning rate was set as 5×10^{-5}. The loss function of mean-squared-error and the AdamW optimizer were adopted. The batch size was set as eight, and the number of epochs was set as 50. The teacher forcing strategy was also used to facilitate the training process with the value of teacher forced samples decays per epoch being 1.5% (Yang and Liu 2020). The training of the ConvLSTM model was implemented using a single NVIDIA RTX 4090 GPU and PyTorch 1.10.

3.2. Model performance and comparison

We first trained several models to evaluate the influence of the kernel size. We tested the kernel sizes of 3×3, 5×5, 7×7, and 9×9, and the results were only marginally different (Supplementary Table S1). To balance the performance and computation costs, we chose the kernel size of 5 because the results were already satisfactory.

The proposed ConvLSTM model was compared against FLUS, LSTM-CA, and CNN-LSTM-CA. Due to the differences of model structure, the data used for calibration for these three models are different. For FLUS, only the 2011 data was used, while for LSTM-CA and CNN-LSTM-CA, the data for the years 2005, 2006, 2007, 2008, 2009, 2010, and 2011 were used. The FoM values for FLUS, LSTM-CA, CNN-LSTM-CA, and the ConvLSTM model are summarized in Table 2. It should be noted that the base year

Table 2. FoM values for the models of FLUS, LSTM-CA, CNN-LSTM-CA and ConvLSTM in the simulations of urban land expansion from 2012 to 2017.

Model	2012	2013	2014	2015	2016	2017
FLUS	0.0390	0.0681	0.0887	0.1075	0.1215	0.1338
LSTM-CA	0.0716	0.1025	0.1169	0.1284	0.1336	0.1421
CNN-LSTM-CA	0.0609	0.1017	0.1258	0.1459	0.1589	0.1646
ConvLSTM	0.1771	0.1985	0.2076	0.2140	0.2186	0.2162

Note: The 2011 urban land map is used as the base year data. For instance, the FoM for year 2016 is the FoM calculated based on the simulated change and the referenced change from 2011 to 2016.

(T_0) for the calculation of FoM is 2011 for all simulations. The results clearly show that FLUS attains smaller FoM values than the other three models, suggesting that without the input of time-series urban land maps, the simulation models are less successful to capture the actual dynamics of urban land expansion. The LSTM-CA model is capable of learning temporal information from time-series urban land maps, and hence achieves greater FoM values than FLUS by approximately 1-3%. CNN-LSTM-CA further improves the FoM values by 1-2% as compared against those of LSTM-CA for the years 2014-2017. The proposed ConvLSTM model, however, achieves the highest FoM values throughout the entire timeline of simulations, which are even 5%-12% more than those of CNN-LSTM-CA. This could be owing to the more advanced architecture of ConvLSTM that can fuse the spatial-temporal information in the same network.

We also carried out additional experiments to test whether the choice of periods for model training and running would affect the performance. In those experiments, we trained the ConvLSTM model with the annual urban land maps for the years 1998-2009 and used it to predict the annual urban land expansion for the subsequent decade (i.e., 2010-2019). The FoM values of the 2010-2019 simulations are provided in Supplementary Table S2. The results suggest that ConvLSTM again significantly outperform FLUS, LSTM-CA, and CNN-LSTM-CA. Therefore, the choice of periods (n) for model training and the number of time steps (m) for prediction do not affect the better performance of the ConvLSTM model, as compared with the other models.

Figure 6 shows the results of map comparison between the referenced and the simulated urban land expansion from 2011 to 2017 for the four models. An evident difference can be observed between the results of ConvLSTM and the others. That is, the results of ConvLSTM demonstrate more mixture of hits (i.e., pixels referenced changed and simulated changed), misses (i.e., pixels referenced changed and simulated unchanged), and false alarms (i.e., pixels referenced unchanged and simulated changed), as highlighted by the cyan dashed circles in Figure 6. These results imply that although the ConvLSTM model cannot always predict the precise locations of change, the predictions are already close to where the referenced change occur. For the other three models, however, the misses and the false alarms are more likely to repel each other. In other words, if the four models are evaluated at coarser scales, better agreements would still be observed in the results of ConvLSTM rather than in those of the other three models.

This is confirmed by Figure 7, which shows the comparison between the amounts of simulated and referenced new urban land during 2012-2017. They were obtained by using the focal sum operation with a moving window of 1.5 km $\times$ 1.5 km. At such a scale, FLUS can explain only 46% of the referenced new urban land, while LSTM-CA and

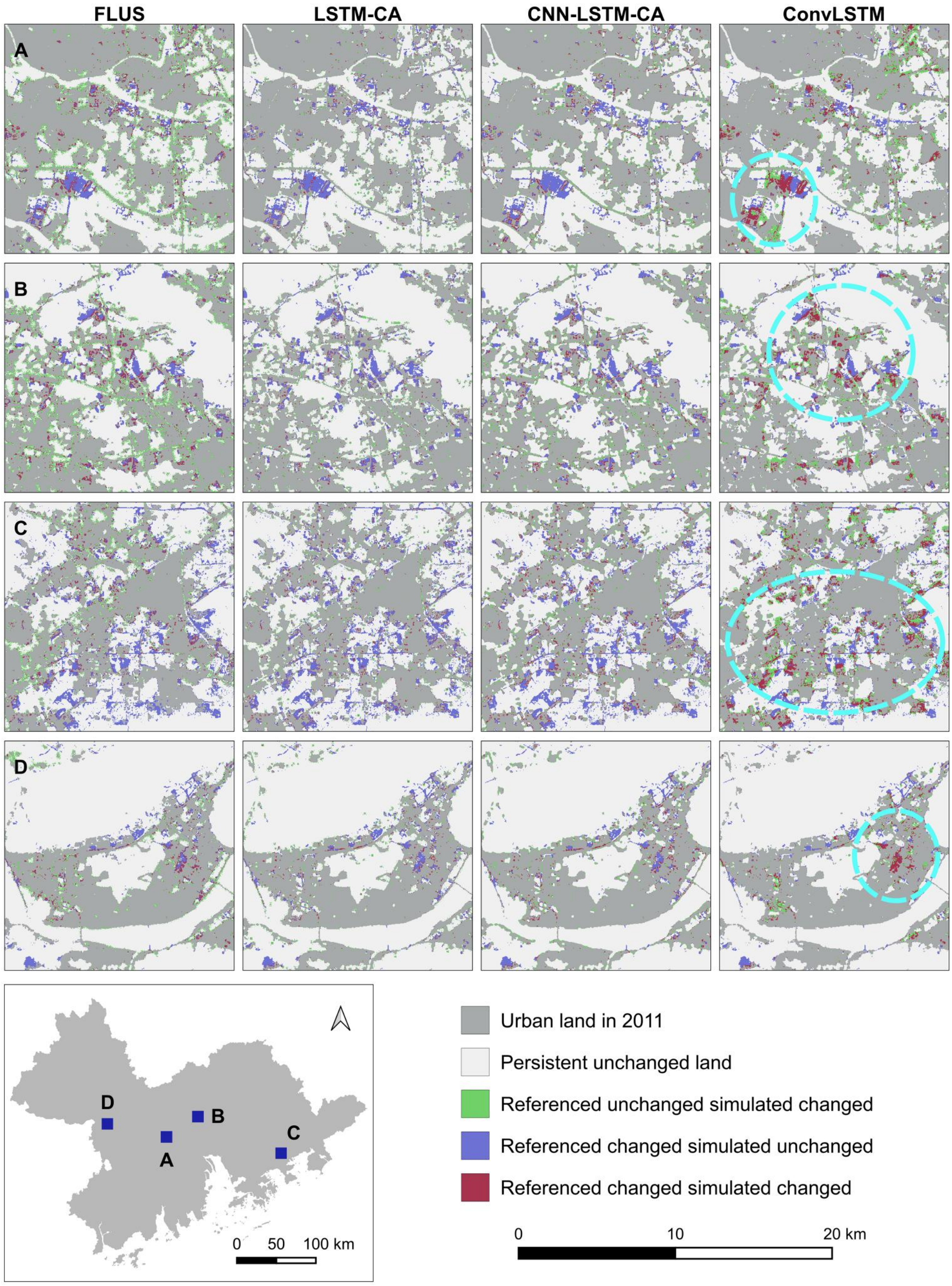

Figure 6. Spatial distributions of hits and errors of the simulated urban land expansion from 2011 to 2017.

CNN-LSTM-CA perform better than FLUS with the explanatory power of 59% and 55%, respectively. For these three models, the histograms clearly demonstrate large disparities between the distributions of the simulated and the referenced new urban land (Figure 7). By contrast, the ConvLSTM model not only has greater explanatory power (70%), but

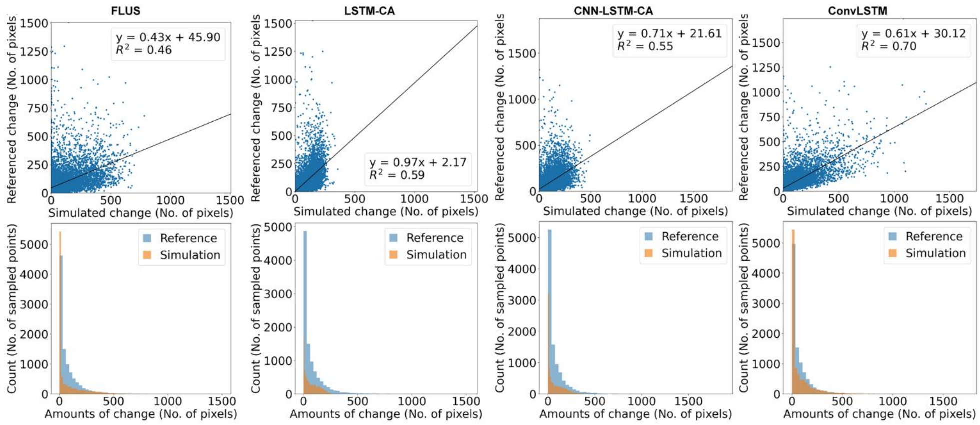

Figure 7. Comparing the simulated and the referenced urban land expansion at the scale of 1.5 km (Sample size = 10000).

also generates a distribution closer to the referenced one, as compared with those yielded by FLUS, LSTM-CA, and CNN-LSTM-CA. Similar analysis was also conducted using the moving windows of 3 km × 3 km (Supplementary Figure S1) and 6 km × 6 km (Supplementary Figure S2). Despite the improvements of R^2 and distribution agreements for all models due to upscaling, the ConvLSTM model (R^2 = 0.80 and 0.87, respectively) still outperforms FLUS (R^2 = 0.61 and 0.71, respectively), LSTM-CA (R^2 = 0.73 and 0.83, respectively), and CNN-LSTM-CA (R^2 = 0.69 and 0.82, respectively).

For the proposed ConvLSTM model, the module of channel attention allows for understanding the importance of each spatial variable in the form of vectors. We exported and transformed the vectors to spatial data, and the results are shown in Supplementary Figure S4. The mean values of importance reflect that D2T (0.71), D2H (0.62), and D2RS (0.61) are the most influential spatial variables, suggesting that proximity to local urban centers and important infrastructure (e.g. highways and railway stations) are the major factors of urban land expansion in the GBA. Slope and D2NR (0.57 and 0.56, respectively) are ranked the 4th and 5th important spatial variables, while D2A and D2C (0.46 and 0.45, respectively) are the least important ones.

3.3. Geographic transferability

Although the proposed ConvLSTM model performs better than the urban CA models, the computation costs of deploying and training ConvLSTM are relatively high. Therefore, for saving the computation costs, it is more desirable if the model trained in one geographic region can be reused in another region. This largely depends on the transferability of the ConvLSTM model. To evaluate the transferability, we conducted several experiments in which the ConvLSTM model is trained using the data of a certain mega-urban region and applied in all three mega-urban regions.

Specifically, the ConvLSTM model is first trained separately using the data of GBA, YRD, and BTT. For convenience, the resulting models are referred to as the GBA model, the YRD model, and the BTT model. These three models are consistently trained using the data of 2000-2011, of which the 2000-2005 data are used as input to simulate the

2012

Applications \ Models	GBA	YRD	BTT
GBA	**0.1771**	0.1748	0.1759
YRD	0.1852	**0.1900**	0.1886
BTT	0.1925	0.1940	**0.1966**

2013

Applications \ Models	GBA	YRD	BTT
GBA	**0.1985**	0.1973	0.1954
YRD	0.2055	**0.2113**	0.2100
BTT	0.2074	0.2109	**0.2124**

2014

Applications \ Models	GBA	YRD	BTT
GBA	**0.2076**	0.2062	0.2029
YRD	0.2169	**0.2239**	0.2219
BTT	0.2147	0.2190	**0.2201**

2015

Applications \ Models	GBA	YRD	BTT
GBA	**0.2140**	0.2131	0.2078
YRD	0.2261	**0.2344**	0.2312
BTT	0.2197	0.2240	**0.2248**

2016

Applications \ Models	GBA	YRD	BTT
GBA	**0.2186**	0.2175	0.2104
YRD	0.2329	**0.2422**	0.2379
BTT	0.2243	0.2281	**0.2295**

2017

Applications \ Models	GBA	YRD	BTT
GBA	**0.2162**	0.2135	0.2053
YRD	0.2343	**0.2412**	0.2371
BTT	0.2239	0.2248	**0.2277**

Figure 8. FoM values of the ConvLSTM models in different experiments.

urban development in the period of 2006-2011. The number of tiles is 6000, with 5000 tiles for training and 1000 tiles for testing. After that the trained models are directly applied to the simulations in all three mega-urban regions. For example, the GBA model is applied to the simulations in both YRD and BTT besides the GBA, and so forth for the YRD model and the BTT model. All simulations are validated using the FoM indicator, of which the results form a 3×3 matrix for every year from 2012 to 2017 (Figure 8).

The results show that the models trained by local data perform slightly better than the ones transferred directly from the other two mega-urban regions. The spatial patterns generated by the three models are also similar (Figure 9). For the same mega-urban region, the simulations yielded by the three models show high agreements, as revealed by Figure 10. For the GBA case, land units (i.e., pixels) where at least two models predict new development account for over 61% of the total land units of predicted change. This proportion is even higher in the cases of YRD and BTT, both of which are 67%. Notably, the proportions of pixels that the three models consistently predict new development are the highest in all three cases, which are 38.0% for GBA, 43.1% for YRD, and 45.9% for BTT. These results suggest that without any fine-tunings or modifications, the proposed ConvLSTM model trained in one mega-urban region still can work reasonably well in other mega-urban regions. This could be related to the similar dynamics of urbanization in the three selected mega-urban regions so that the information learned in one mega-urban region can be immediately used in the others.

3.4. Ablation experiments

Ablation experiments were conducted to demonstrate the effectiveness of the components used to form the proposed ConvLSTM model. The experiments used all three

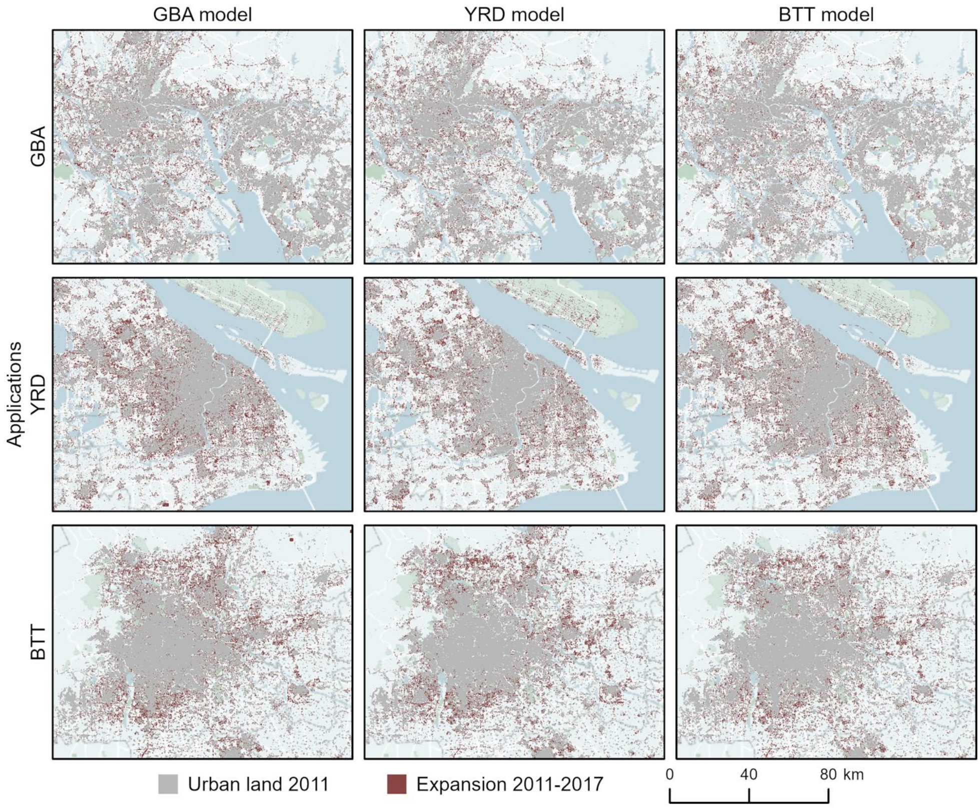

Figure 9. Representative examples of the urban land expansion simulations from 2011 to 2017 generated by the transferred ConvLSTM models.

sets of urban data for GBA, YRD, and BTT to compare the performance of four models. They include the basic model with only the vanilla ConvLSTM cells (M_{Van}), the basic model combined with the component of contextual embedding (M_{Van+CE}), the basic model combined with the component of channel attention ($M_{Van+ChA}$), and the complete form of the proposed ConvLSTM model ($M_{complete}$). Table 3 summarizes the FoM values for the simulation of urban land expansion from 2012 to 2017, yielded by the four models under comparison. The results suggest that complementing the basic model with the component of contextual embedding consistently improves the simulation accuracy in all three cases of experiments. By contrast, the influence of adding only the component of channel attention to the basic model varies across different experiments, in which the simulation accuracy reduces in the case of BTT. If both the components of contextual embedding and channel attention are added, however, the simulation accuracy can further improve as compared with M_{Van+CE}. Therefore, the ablation experiments confirm the effectiveness of the proposed ConvLSTM model.

4. Discussion

The results of the case study in the GBA reveal the significant improvements of simulation accuracy for models that are capable of learning spatial-temporal information

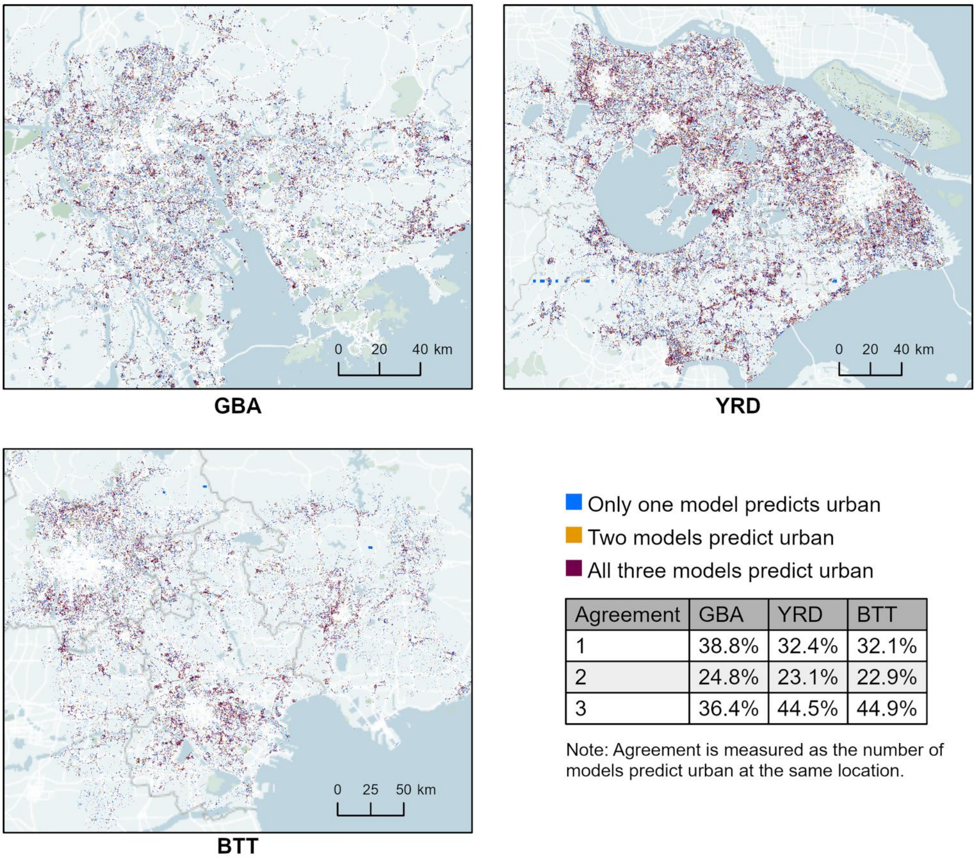

Figure 10. Agreement among the simulations yielded by the GBA model, the YRD model, and the BTT model.

Table 3. FoM values of the ablation experiments using the GBA, YRD, and BTT data.

Region	Year	M_{Van}	M_{Van+CE}	$M_{Van+ChA}$	$M_{complete}$
GBA	2012	0.1697	0.1752	0.1714	0.1771
	2013	0.1944	0.1960	0.1945	0.1985
	2014	0.2037	0.2046	0.2032	0.2076
	2015	0.2087	0.2095	0.2114	0.2140
	2016	0.2124	0.2127	0.2157	0.2186
	2017	0.2088	0.2105	0.2128	0.2162
YRD	2012	0.1821	0.1880	0.1801	0.1900
	2013	0.2055	0.2091	0.2053	0.2113
	2014	0.2166	0.2211	0.2183	0.2239
	2015	0.2241	0.2310	0.2285	0.2344
	2016	0.2294	0.2377	0.2363	0.2422
	2017	0.2286	0.2355	0.2384	0.2412
BTT	2012	0.1855	0.1934	0.1878	0.1966
	2013	0.2066	0.2101	0.2060	0.2124
	2014	0.2156	0.2181	0.2141	0.2201
	2015	0.2207	0.2228	0.2187	0.2248
	2016	0.2250	0.2268	0.2238	0.2295
	2017	0.2217	0.2244	0.2228	0.2277

from time-series urban land maps, such as LSTM-CA, CNN-LSTM-CA, and ConvLSTM. Compared with CNN-LSTM-CA, which uses two separating networks of CNN and LSTM to learn separately the spatial and temporal information from data, the ConvLSTM model embeds the convolution operation to each of the vanilla ConvLSTM cells and hence can simultaneously obtain the spatial and temporal information from data in the same network. The results of model comparison clearly show that ConvLSTM significantly outperforms CNN-LSTM-CA by 5%-12% in terms of FoM (Table 2), confirming the effectiveness of the proposed architecture.

Moreover, the proposed ConvLSTM presents a distinct way for modeling urban land expansion. Conventional CA-based urban simulation models are developed based on the idea of self-organizing, and operate by iterating the constant local rules that specify how neighborhood interaction and spatial variables affect urban land change (Tobler 1979, Phipps 1989). An important issue in conventional CA-based models is how to align the iterations to the actual time. Most of the existing studies, if not all, use a compromised representation of temporal dimension with a fixed value of iteration count, which is defined according to some constraints such as land demand. Due to such a technical implementation, the effect of path-dependence arises and becomes one of the sources of simulation errors (Brown *et al.* 2005).

Additionally, in conventional CA-based urban simulation models, neighborhood interaction is often represented in the form of development density (i.e., counting the number of land units that are developed within a given neighborhood), and the neighboring land units are either equally weighted or weighted by the distance from the neighborhood center (or its variants) (Wang *et al.* 2021, Roodposhti *et al.* 2020, Liao *et al.* 2016, Zhang *et al.* 2023a). It is questionable to use the traditional one-fits-all density calculation to represent the neighborhood interaction in different spatial contexts.

For the proposed ConvLSTM model, however, the temporal dimension is explicit and matches the actual timeline. In this study, we had used ConvLSTM to process the annual urban land maps and make prediction annually. Building upon this feature, the spatial-temporal relationships between urban land data can be learnt and implicitly represented by ConvLSTM. In particular, the multiple convolution operations in ConvLSTM have two major advantages over traditional representation of neighborhood interaction: First, compared with the traditional uniform calculation of neighborhood development density, the multiple convolution operations are more advanced because it can inexplicitly capture the different aspects of neighborhood interaction in addition to development density. Second, the contributions of neighbors (i.e., the weights of the kernels; see Section 2.2.4) are directly learnt from time-series urban land maps and hence more reliable than using the traditional equal weights or distance-adjusted weights that are manually defined a priori. Nevertheless, compared with the traditional density calculation, the multiple convolutions are less intuitive to understand how neighborhood interaction affects urban development.

Another notable feature of the ConvLSTM model is its promising transferability among the selected mega-urban regions, as shown by Figures 8-10. Nevertheless, the examination of transferability for the ConvLSTM model is still a preliminary analysis at the current stage. More cases of modeling should be considered in future study, particularly those small- and medium-size urbanized areas in different countries. Moreover, the proposed ConvLSTM model still cannot effectively addressed the issue of enclave simulations (see Supplementary Figure S3), which is important for modeling urban growth in fast developing regions. The integration of ConvLSTM and a patch-based simulation approach (Chen *et al.* 2019a, Yang *et al.* 2023) perhaps can alleviate this problem, which can be fulfilled in future studies. Finally, it is feasible to use the ConvLSTM model to provide finer-resolution (e.g. 30 m) projections of future urban development at the global scale, thereby facilitating the understanding of global change and its impacts.

5. Conclusions

This study proposes a novel ConvLSTM model for the simulation of urban land expansion based on the integration of channel attention, contextual embedding, and vanilla ConvLSTM cells. Compared with conventional CA-based models, the proposed model is more advanced in that it can better leverage the open access annual urban land maps to learn simultaneously the spatial structure and the temporal dependency of historical urban development, and further generate the annual predictions of future development for multiple years, i.e., a Maps-to-Maps manner.

The results of the case study in the GBA show that the ConvLSTM model outperforms several state-of-the-art urban CA models, including FLUS, LSTM-CA, and CNN-LSTM-CA. Compared with these models, ConvLSTM features the highest agreements at the pixel-level and the grid-levels with the resolutions ranging from 1.5 km to 6 km. Furthermore, the ConvLSTM model also shows satisfactory transferability. The experiments in the mega-urban regions of GBA, YRD, and BTT indicate that the ConvLSTM model trained in one mega-urban region can be conveniently applied in the other two without further tuning. Future studies will enhance the examination of the model's transferability by increasing the diversity of case study areas (e.g. urban areas with different sizes and geographic contexts), and extend the applications of the model to wider fields such as global change.

Acknowledgments

The authors are grateful to the associate editor, Song Gao, and the anonymous reviewers for their valuable comments and suggestions.

Authors contributions

Zihao Zhou: Methodology, formal analysis, validation, writing—original draft preparation; Yimin Chen: Conceptualization, supervision, project administration, funding acquisition, writing—review and editing; Xiaoping Liu: Conceptualization, writing—review and editing; Xinchang Zhang: Investigation, writing—original draft preparation; Honghui Zhang: Investigation, writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was funded by National Natural Science Foundation of China (Grant No. 42271415 and 42322110) and Guangdong Natural Science Funds for Distinguished Young Scholar (Grant No. 2021B1515020104).

Notes on contributors

Zihao Zhou is a master's student of GIScience at Sun Yat-sen University, Guangzhou, China. His research interests include urban simulation and urban big data analysis.

Yimin Chen is an Associate Professor of GIScience at Sun Yat-sen University, Guangzhou, China. His research interests include machine learning and urban applications.

Xiaoping Liu is a Professor of GIScience at Sun Yat-sen University, Guangzhou, China. His research interests include land use change models and applications.

Xinchang Zhang is a Professor of GIScience at Guangzhou University, Guangzhou, China. His research interests include spatial data analysis and GISystem development.

Honghui Zhang is the Co-President of Guangdong Guodi Planning Science Technology Co., Ltd, China. His research interests include urban big data analysis and urban planning.

Data and codes availability statement

The materials that support the findings of this study are available via https://doi.org/10.6084/m9.figshare.23918658. The code can also be accessed via https://github.com/kwtk86/UrbanM2M.

References

Addae, B., and Dragićević, S., 2022. Integrating multi-criteria analysis and spherical cellular automata approach for modelling global urban land-use change. *Geocarto International*. doi: 10.1080/10106049.2022.2152498.

Aslam, R.W., Shu, H., and Yaseen, A., 2023. Monitoring the population change and urban growth of four major Pakistan cities through spatial analysis of open source data. *Annals of GIS*, 29 (3), 355–367.

Brown, D.G., *et al.*, 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19 (2), 153–174.

Cai, J., and Chen, Y., 2021. MHA-Net: Multipath hybrid attention network for building footprint extraction from high-resolution remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 5807–5817.

Cao, M., *et al.*, 2019. A grey wolf optimizer–cellular automata integrated model for urban growth simulation and optimization. *Transactions in GIS*, 23 (4), 672–687.

Chai, Z., *et al.*, 2022. CMS-LSTM: Context embedding and multi-scale spatiotemporal expression LSTM for predictive learning. *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 01–06.

Chatterjee, U., and Majumdar, S., 2022. Impact of land use change and rapid urbanization on urban heat island in Kolkata city: A remote sensing based perspective. *Journal of Urban Management*, 11 (1), 59–71.

Chen, G., *et al.*, 2020. Global projections of future urban land expansion under shared socioeconomic pathways. *Nature Communications*, 11 (1), 537.
Chen, Y., 2022. An extended patch-based cellular automaton to simulate horizontal and vertical urban growth under the shared socioeconomic pathways. *Computers, Environment and Urban Systems*, 91, 101727.
Chen, Y., *et al.*, 2016. Capturing the varying effects of driving forces over time for the simulation of urban growth by using survival analysis and cellular automata. *Landscape and Urban Planning*, 152, 59–71.
Chen, Y., *et al.*, 2019a. Simulating urban growth boundaries using a patch-based cellular automaton with economic and ecological constraints. *International Journal of Geographical Information Science*, 33 (1), 55–80.
Chen, Y., *et al.*, 2019b. Tele-connecting China's future urban growth to impacts on ecosystem services under the shared socioeconomic pathways. *The Science of the Total Environment*, 652, 765–779.
D'Acci, L., and Batty, M., 2019. *The mathematics of urban morphology*. Cham, Switzerland: Springer.
Feng, Y., and 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 (1), 74–97.
Gao, C., *et al.*, 2020. Modeling urban growth using spatially heterogeneous cellular automata models: Comparison of spatial lag, spatial error and GWR. *Computers, Environment and Urban Systems*, 81, 101459.
Gao, J., and O'Neill, B.C., 2020. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nature Communications*, 11 (1), 2302.
Geng, J., *et al.*, 2022. A hybrid spatiotemporal convolution-based cellular automata model (ST-CA) for land-use/cover change simulation. *International Journal of Applied Earth Observation and Geoinformation*, 110, 102789.
Gong, P., *et al.*, 2020. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment*, 236, 111510.
Graves, A., 2012. Long short-term memory. In: Graves, A. ed. *Supervised sequence labelling with recurrent neural networks*. Heidelberg, Germany: Springer, 37–45.
Guan, X., *et al.*, 2023. HGAT-VCA: Integrating high-order graph attention network with vector cellular automata for urban growth simulation. *Computers, Environment and Urban Systems*, 99, 101900.
Hagen-Zanker, A., 2006. Map comparison methods that simultaneously address overlap and structure. *Journal of Geographical Systems*, 8 (2), 165–185.
Hawker, L., *et al.*, 2022. A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17 (2), 024016.
Huang, X., *et al.*, 2022. Toward accurate mapping of 30-m time-series global impervious surface area (GISA). *International Journal of Applied Earth Observation and Geoinformation*, 109, 102787.
Li, X., *et al.*, 2021. Critical role of temporal contexts in evaluating urban cellular automata models. *GIScience & Remote Sensing*, 58 (6), 799–811.
Liao, J., *et al.*, 2016. Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. *Environmental Modelling & Software*, 75, 163–175.
Liu, J., *et al.*, 2021. Simulation of dynamic urban expansion under ecological constraints using a long short term memory network model and cellular automata. *Remote Sensing*, 13 (8), 1499.
Liu, X., *et al.*, 2017. A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*, 168, 94–116.
Liu, X., *et al.*, 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability*, 3 (7), 564–570.
Li, X., Zhou, Y., and Chen, W., 2020. An improved urban cellular automata model by using the trend-adjusted neighborhood. *Ecological Processes*, 9 (1), 13.

Mahtta, R., *et al.*, 2022. Urban land expansion: The role of population and economic growth for 300+ cities. *Npj Urban Sustainability*, 2 (1), 5.

Masrur, A., and Yu, M., 2023. Spatiotemporal attention ConvLSTM networks for predicting and physically interpreting wildfire spread. *Artificial Intelligence in Earth Science*. Amsterdam, Netherlands: Elsevier, 119–156.

Meentemeyer, R.K., *et al.*, 2013. FUTURES: Multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers*, 103 (4), 785–807.

Melis, G., Kočiský, T., and Blunsom, P., 2019. Mogrifier LSTM. *arXiv preprint arXiv*, 1909.01792.

Merity, S., Keskar, N.S., and Socher, R., 2018. Regularizing and optimizing LSTM language models. *International Conference on Learning Representations*.

Modiri, M., Gholami, Y., and Hosseini, S.A., 2023. Urban growth dynamics modeling through urban DNA in Tehran metropolitan region. *Annals of GIS*, 29 (1), 55–74.

Moghar, A., and Hamiche, M., 2020. Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, 170, 1168–1173.

Phipps, M., 1989. Dynamical behavior of cellular automata under the constraint of neighborhood coherence. *Geographical Analysis*, 21 (3), 197–215.

Pijanowski, B.C., *et al.*, 2014. A big data urban growth simulation at a national scale: configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. *Environmental Modelling & Software*, 51, 250–268.

Pontius Jr, R.G., 2018. Criteria to confirm models that simulate deforestation and carbon disturbance. *Land*, 7 (3), 105.

Qian, Y., *et al.*, 2020. Coupling cellular automata with area partitioning and spatiotemporal convolution for dynamic land use change simulation. *The Science of the Total Environment*, 722, 137738.

Rienow, A., *et al.*, 2021. Modeling urban regions: Comparing random forest and support vector machines for cellular automata. *Transactions in GIS*, 25 (3), 1625–1645.

Roodposhti, M.S., Hewitt, R.J., and 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., *et al.*, 2021. Integrating a forward feature selection algorithm, random forest, and cellular automata to extrapolate urban growth in the Tehran-Karaj region of Iran. *Computers, Environment and Urban Systems*, 87, 101595.

Shi, X., *et al.*, 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.

Simkin, R.D., *et al.*, 2022. Biodiversity impacts and conservation implications of urban land expansion projected to 2050. *Proceedings of the National Academy of Sciences of the United States of America*, 119 (12), e2117297119.

Staudemeyer, R.C., and Morris, E.R., 2019. Understanding LSTM – A tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv*, 1909.09586.

Sumbo, D.K., Anane, G.K., and Inkoom, D.K.B., 2023. Peri-urbanisation and loss of arable land': Indigenes' farmland access challenges and adaptation strategies in Kumasi and Wa, Ghana. *Land Use Policy*, 126, 106534.

Tobler, W.R., 1979. Cellular geography. *Philosophy in Geography*, 20, 379–386.

Tong, X., and Feng, Y., 2020. A review of assessment methods for cellular automata models of land-use change and urban growth. *International Journal of Geographical Information Science*, 34 (5), 866–898.

Wang, H., *et al.*, 2021. Simulating urban land growth by incorporating historical information into a cellular automata model. *Landscape and Urban Planning*, 214, 104168.

Wang, J., *et al.*, 2022. Simulating large-scale urban land-use patterns and dynamics using the U-Net deep learning architecture. *Computers, Environment and Urban Systems*, 97, 101855.

Woo, S., *et al.*, 2018. Cbam: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19.

Xing, W., *et al.*, 2020. A novel cellular automata model integrated with deep learning for dynamic spatio-temporal land use change simulation. *Computers & Geosciences*, 137, 104430.

Yang, J., *et al.*, 2023. Simulating urban expansion using cellular automata model with spatiotemporally explicit representation of urban demand. *Landscape and Urban Planning*, 231, 104640.

Yang, Z., and Liu, C., 2020. 2D map estimation via teacher-forcing unsupervised learning. *2020 International Conference on Connected and Autonomous Driving (MetroCAD)*, 91–97.

Yu, Y., *et al.*, 2019. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, 31 (7), 1235–1270.

Zhai, Y., *et al.*, 2020. Simulating urban land use change by integrating a convolutional neural network with vector-based cellular automata. *International Journal of Geographical Information Science*, 34 (7), 1475–1499.

Zhang, B., *et al.*, 2023a. A size-adaptive strategy to characterize spatially heterogeneous neighborhood effects in cellular automata simulation of urban growth. *Landscape and Urban Planning*, 229, 104604.

Zhang, J., *et al.*, 2023b. Modelling urban expansion with cellular automata supported by urban growth intensity over time. *Annals of GIS*, 29 (3), 337–353.

Zhou, Y., *et al.*, 2018. A global record of annual urban dynamics (1992–2013) from nighttime lights. *Remote Sensing of Environment*, 219, 206–220.

Zhou, Y., *et al.*, 2023. A novel spatio-temporal cellular automata model coupling partitioning with CNN-LSTM to urban land change simulation. *Ecological Modelling*, 482, 110394.