
Volume 130
30 November 2013
ISSN 0301-4797
ELSEVIER
Journal of
Environmental
Management

Journal of Environmental Management 130 (2013) 106–116

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Early warning of illegal development for protected areas by integrating cellular automata with neural networks

Xia Li*, Chunhua Lao, Yilun Liu, Xiaoping Liu, Yimin Chen, Shaoying Li, Bing Ai, Zijian He

Guangdong Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, 135 West Xingang RD., Guangzhou 510275, PR China

ARTICLE INFO

Article history:
Received 11 November 2012
Received in revised form
22 August 2013
Accepted 27 August 2013
Available online 24 September 2013

Keywords:
Early warning systems
Cellular automata
Neural networks
Illegal development
Geographical information systems

ABSTRACT

Ecological security has become a major issue under fast urbanization in China. As the first two cities in this country, Shenzhen and Dongguan issued the ordinance of Eco-designated Line of Control (ELC) to "wire" ecologically important areas for strict protection in 2005 and 2009 respectively. Early warning systems (EWS) are a useful tool for assisting the implementation ELC. In this study, a multi-model approach is proposed for the early warning of illegal development by integrating cellular automata (CA) and artificial neural networks (ANN). The objective is to prevent the ecological risks or catastrophe caused by such development at an early stage. The integrated model is calibrated by using the empirical information from both remote sensing and handheld GPS (global positioning systems). The *MAR* indicator which is the ratio of missing alarms to all the warnings is proposed for better assessment of the model performance. It is found that the fast urban development has caused significant threats to natural-area protection in the study area. The integration of CA, ANN and GPS provides a powerful tool for describing and predicting illegal development which is in highly non-linear and fragmented forms. The comparison shows that this multi-model approach has much better performances than the single-model approach for the early warning. Compared with the single models of CA and ANN, this integrated multi-model can improve the value of *MAR* by 65.48% and 5.17% respectively.

1. Introduction

Urban expansion has become a global phenomenon which is at the cost of losing ecological and agricultural land. Under the pressure of protecting natural heritage, more than 12,700 protected areas (e.g. parks and wildlife refuges) have been established around the world, accounting for 8.8% of the Earth's land surface (Liu et al., 2001; McDonald et al., 2001). In many developed counties, there is a long history of conserving protected areas (Dompka, 1996). However, it is rather a challenge task for China to protect its treasured land resources under the pressure of rapid urban expansion since the economic reform in 1978. Many fast growing cities in China have to encroach on ecological or agricultural land for keeping the economy growing. In China, a city will expand by 3 per cent on average if its economy, measured by gross domestic product, grows by 10 per cent according to statistical analysis (Deng et al., 2008).

Chinese economic development is accompanied by many environmental and ecological problems, such as the degradation of rural, agrarian, and ecological systems (Forman, 2008; Yeh and Li, 1999). Ecological security has become a major concern for this fast growing country. Actually, the concept of ecological security was first proposed by the governments of the United States for tackling environmental problems during urban growth (Ezeonu and Ezeonu, 2000). Ecological security may mean the safety from injury, harm, or danger without damage to the natural systems. As a tool to implement the concept of ecological security, early warning systems (EWS) are used to recognize the "early warning" signs of environmental or ecological degradation. EWS is to facilitate the systematic collection of information which can shed light on the causes and dynamics of natural calamities. Many EWS actually involve the techniques of realtime multiple-source data collection, data transmission, evaluation and analysis for timely dissemination of early warning (Quansah et al., 2010). Advances in remote sensing and GPS techniques have resulted in more reliable, high frequency and automated collection of critical ecological and environmental status.

Besides data collection, evaluation and analysis methods are also important for the construction of EWS. EWS should have the

* Corresponding author. Tel./fax: +86 (0)20 84115833.
E-mail addresses: lixia@mail.sysu.edu.cn, lixia@graduate.hku.hk (X. Li).
URL: http://www.geosimulation.cn/

0301-4797/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jenvman.2013.08.055

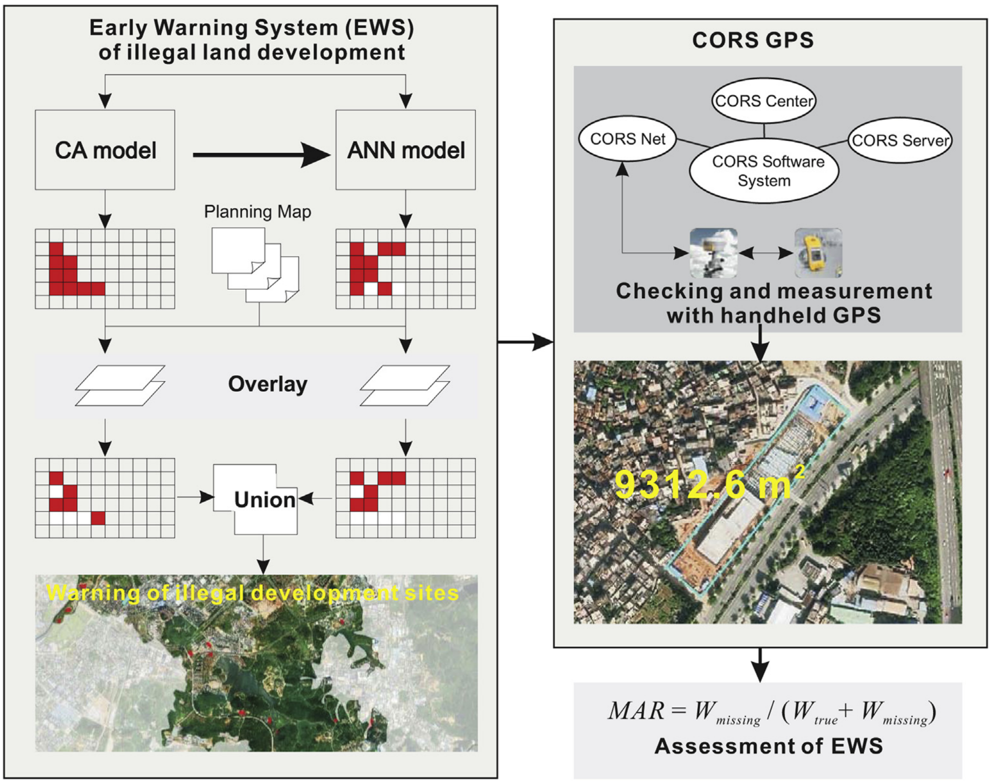

Fig. 1. Early Warning System (EWS) for predicting illegal development by integrating cellular automata (CA) and neural networks (ANN).

capability of modeling the dynamics of biological and environmental elements which are involved for the assessment of ecological security (Tegler et al., 2001; Barlindhaug et al., 2007; Hockey and Curtis, 2009; Li et al., 2010). There are quite diverse techniques for identification, description, evaluation and prediction of ecological security. Li et al. (2010) proposed an index system for landscape ecological security (LES) using three dimensions, six factors, and three weights. There are few studies on the development of early warning mechanisms for preventing illegal development in fast growing regions. New tools are needed for effectively predicting and assessing the potential impacts of illegal development within protected natural areas. Gong et al. (2009) first proposed a method to assess ecological security by using cellular automata (CA).

Illegal development which usually occurs during the period of rapid urban growth refers to the construction without permits and authorized blueprints. Government reports and academic studies have revealed that illegal development is extensive and persistent in China since the economic reform in 1978. There were already 1.84 million cases or 1730 km^2 of illegal development in the mid-1980s, and another 1 million cases or 2000 km^2 of such development in the 1990s (Tang and Chung, 2002). This situation persisted in the 2000s because of huge land demand. A total of 13, 000 cases or 160 km^2 of illegal development was further identified for the period of October 2005–October 2006 according to the 7th national land use investigation by using remote sensing techniques (Wang and Li, 2008). Their studies indicated that the portion of illegal development amounted to 51% (cases) and 22% (area) of the total development. Moreover, there were eight administrative cities with more than 80% of their total amount of development identified as illegal. This type of unprecedented growth has resulted in serious urban sprawl in some large cities, such as Beijing (Wong and Tang, 2005) and Guangzhou (Wu and Yeh, 1999).

Various attempts have been made in China to protect its shrinking agricultural land resources during urbanization. These attempts include the implementation of development containment strategies to limit city size (both in terms of population and the built-up area), the restriction of new development in important agricultural areas, the promotion of high-density development, and the designation of greenbelts (Zhao, 2011). Particularly, strict legislation can be imposed to contain urban development within a reasonable boundary. Shenzhen and Dongguan are the first two cities to promulgate the ordinance of Eco-designated Line of Control (ELC) which is to “wire” important ecological land for legal protection. Shenzhen is the first administrative city to adopt the act of ecological land protection in China. On 1 November 2005, the first law in China for protecting important ecological land from urban development was issued by Act 145 of Shenzhen’s Government. In this Act, a total of 974 km^2 (about 49.9% of its total administrative land area) has been delineated as the protected land. However, it is found that illegal development is quite common within ELC. Li (2007) reported that the average approved amount of development was about 20 km^2 annually, but the actual amount of development was as high as 48 km^2. Moreover, only 43% of the existing urban land (724 km^2) in 2007 had been legally approved by the government agencies.

It should be too late to obtain the information of illegal development from field investigations or remote sensing monitoring for the management purpose. Instead, early warning of possible illegal development is appealing for preventing the ecological risks or catastrophe before they could take place. However, the early warning involves complex factors and uncertainties which are difficult to represent and handle by using top-down mathematical equations. This problem can be partially solved by using bottom-up simulation methods, such as cellular automata and agent-based models. These models are just implemented based on some local interaction rules (Li et al., 2011). Recent years have witnessed the fast development of various bottom-up simulation models for predicting land use change trajectories and exploring possible development options. Particularly, a family of cellular automata (CA) have been developed, including SLEUTH (Clarke et al., 1997), GeoSOS (Li et al., 2011), and CLUE-S (Verburg et al., 2002). Although there is a growing trend of using agent-based models, CA have been considered to be convenient and well defined in terms of model structures and model calibration (Li et al., 2011).

Few studies have been carried out to use these models for early warning of ecological and environmental risks caused by illegal land development. This paper will present a multi-model approach by integrating CA and artificial neural networks (ANN) for improving the predictability of EWS. CA will be used to simulate the illegal development for future years. ANN are also incorporated to improve the simulation performance because the development is an extremely nonlinear process. Moreover, handheld GPS with high-precision will be utilized to provide empirical information to train and validate the EWS. This integrated model is expected to provide a useful tool for researchers and policy-makers to understand and predict illegal development under different future-oriented environmental policies. This model will also yield timely guidance in deciding mitigation actions against illegal development for fast growing regions.

2. Early warning of illegal development by integrating cellular automata and neural networks

As a special type of land use conversion, illegal land development is subjected to a series of uncertainties and characterized by a diversity of scales and processes. This may explain why illegal development is usually in fragmented or leap-frog patterns (Zhao et al., 2009). In this study, a multi-model approach is proposed to capture the characteristics of complex land development for implementing the early warning system (EWS). This multi-model approach is developed in three-folds: 1) simulating urban expansion by using a process model, a logistic cellular automaton (Logistic-CA); 2) capturing highly non-linear feature of illegal development by using a neutral network (ANN); and 3) integrating these two models for improving the performances of EWS.

The integration of CA and ANN is to provide the complementary information which is crucial for the early warning of illegal development (Fig. 1). A logistic-CA is directly used to simulate the patterns and processes of urban development. ANN is also developed to improve the accuracy of warning illegal development. Since the ANN model is only calibrated by using one year of high-resolution data, it cannot directly simulate illegal development for future years. Instead, a logistic-CA which is a process model will be utilized to obtain the total amounts of predicted illegal development for various future years. These amounts of illegal development are then treated as the constraints of ANN so that this model can predict the patterns of illegal development for these years. The combination of these two models is to improve the accuracy of identifying potential illegal development for the EWS. The following section will describe the details of the proposed methodology (Fig. 1).

2.1. Logistic cellular automaton

For simulating urban dynamics and land use changes, CA need to be parameterized with statistical or artificial intelligent techniques by using empirical information (Li et al., 2011). The parameterization of CA based on logistic regression has proven to be effective for quantifying the potential interactions related to land use dynamics (Wu, 2002; Li et al., 2008; Lin et al., 2011). Before the parameterization, it is essential to identify the driving factors which determine land use dynamics for constructing CA. Explanatory variables for land use dynamics usually include the accessibility to the built and natural amenity features, and the site and neighborhood properties (Conway and Wellen, 2011). The accessibility can be quantified by the proximities to various attraction centers, such as transport, roads, railways, urban centers and commercial centers (Wu and Webster, 1998; Li and Yeh, 2002; Lin et al., 2011).

The logistic-CA is based on the estimation of the conversion probability from a series of spatial variables (Wu, 2002; Li et al., 2008):

$$p_{ij}^t = \frac{\exp\left(z_{ij}^t\right)}{1+\exp\left(z_{ij}^t\right)} = \frac{1}{1+\exp\left(-z_{ij}^t\right)} \tag{1}$$

where p_{ij}^t is the conversion probability at time t for cell ij; $z_{ij}^t = a_0 + a_1x_1 + a_2x_2 + \cdots + a_mx_m + \cdots + a_Mx_M$, a_0 is the constant, x_m is a spatial (physical) variable representing a driving force for urban development, and a_m is the parameter (weight) of variable x_m.

The above equation only addresses the global interactions which are in a function of various proximity variables for land use conversion. Actually, local (neighborhood) interactions between different land use types compose the important part of transition rules of CA. Local interactions are related to site and neighborhood properties. Moreover, some geographical constraints (e.g. topography, protected ecological land and planning schemes) should be included to address environmental and ecological conditions. By considering all these factors, the development probability is further revised as follows (Li et al., 2008, 2011):

$$p_{ij}^t = \left(1+(-\ln\gamma)^{\alpha}\right)\frac{1}{1+\exp\left(-z_{ij}^t\right)} \times f\left(\Omega_{ij}^t\right) \times \mathrm{con}\left(s_{ij}^t\right) \tag{2}$$

where γ is a stochastic factor ranging from 0 to 1, α is a parameter to control the stochastic degree, $f(\Omega_{ij}^t)$ is the development intensity in the neighborhood of Ω_{ij}, and $\mathrm{con}(s_{ij}^t)$ is the combined constraint score ranging from 0 to 1.

At each iteration of simulation, p_{ij}^t is compared with a threshold value to determine if a non-urbanized cell will be converted into an urbanized cell:

$$S_{ij}^{t+1} = \begin{cases} \text{Converted}, & p_{ij}^t \geq \gamma \\ \text{Non Converted}, & p_{ij}^t < \gamma \end{cases} \tag{3}$$

where γ is a threshold value.

The threshold (γ) is determined according to the actual land demand which is usually obtained from observation data or an exogenous growth model. For example, this value can be estimated in such way that the total number of converted cells will be equal to the actual one observed from classified remote sensing data (Li and Yeh, 2002). The implementation of logistic-CA is simple as it has been widely used in many studies (Wu, 2002; Li et al., 2008). Actually, it is available in the free package of GeoSOS (Li et al., 2011).

2.2. Artificial neural networks

Artificial neural networks (ANN) are developed by simulating human's learning and recalling abilities. ANN have been used to solve many practical problems, such as pattern classification, complexity and dimension reduction, and temporal prediction (Grossberg, 1988; Chen and Billings, 1992). These models have also been applied to the analysis and modeling of various geographical problems (Openshaw, 1998). Studies indicate that ANN can well deal with complex nonlinear relationship between the driving variables and land use dynamics (Li and Yeh, 2002). It is found that ANN even have higher overall accuracy and kappa statistics than other models (e.g. logistic regression) for modeling land use changes (Lin et al., 2011). This is the main reason that ANN are incorporated in this EWS for predicting illegal development.

For modeling a complex non-linear mapping problem, neural networks usually contain three layers of neurons: input, hidden and output layers. In the input layer, each neuron which is associated with an input variable (e.g. the independent variable for land use conversion). The output layer yields the conversion probability of land use changes (Li and Yeh, 2002).

ANN are usually trained based on a back-propagation learning mechanism (Foody, 1996). This mechanism iteratively minimizes an error function over the network (calculated) outputs and desired (known) outputs. After the optimized weights have been obtained, this parameterized three-layer network can be used to predict the conversion probability for simulating illegal development.

2.3. Early warning of illegal development based on a multi-model approach

The early warning system (EWS) is formulated by using a number of techniques, such as remote sensing, cellular automata, neural networks, and GPS. Actually, this system consists of a number of functions, such as remote sensing monitoring, handheld GPS checking, estimation of development probabilities, assessment for warning, and prediction of perceived potential alarms (Fig. 1). In order to reduce the possibility of missing warnings, we use a multi-model approach to identify the potential sites of illegal development for future years. All these sites identified by either of these two models will be treated as warnings. This is based on the "OR" operator for combining CA and ANN simulation results. Such operator may create a fair amount of over-warnings. However, it is expected that this approach can reduce the risk of missing warnings.

As an important part of the EWS, handheld GPS are used to confirm, delineate and label the warnings on the ground. The field measurement is based on a Continuous Operational Reference System (CORS) which uses local reference stations to provide up to centimeter-level accuracy. The high-precision GPS will obtain detailed geometric information (e.g. shape and area) for each confirmed site of illegal development. The ground-true information is important for training and validating EWS.

The assessment of EWS is usually carried out by using the *FAR* indicator which is the ratio of false alarms, or unverified warnings, to all the warnings issued (Barnes et al., 2007):

$$FAR = W_{\text{false}} / \left(W_{\text{true}} + W_{\text{false}} \right) \tag{4}$$

where W_{true} and W_{false} are the area of true (confirmed) and false warnings respectively based on field investigations.

The perception of warning accuracy is important for implementing early warning systems. A warning may turn out to be wrong because of the complexity of natural phenomena. In practice, a larger area of warnings should be identified because of a safety reason (Barnes et al., 2007). There is not the so-called "*cry-wolf effect*". The 'do no harm' principle will favor over-warnings since a fair amount of false-warnings can be removed by further field investigation. For example, the public still needs the forecasting or has the confidence in the forecasting if 70% of tornado warnings turned out to be false.

In this study, the *MAR* indicator which is the ratio of missing-warnings to all the warnings is thus proposed for the assessment:

$$MAR = W_{\text{missing}} / \left(W_{\text{true}} + W_{\text{missing}} \right) \tag{5}$$

where W_{true} and W_{missing} are the area of true (confirmed) and missing warnings respectively based on field investigations.

3. Model implementation and results

3.1. The study area

Dongguan is a fast growing city in the Pearl River Delta, China. Like many cities in this region, the administrative area of Dongguan consists of the urban districts and many rural towns. Actually, it has four urban districts and 29 towns with a total area of 2465 km^2. Situated just about 100 km north of Hong Kong, this administrative city is attractive for manufacturing industry because of its strategic geographical location. It used to be an agricultural county before 1986, but now has been converted to one of the largest electronic manufacturing centers in the world after about three decades of rapid industrialization (Li et al., 2011).

The fast urbanization and industrialization in this region have caused a lot of land use problems, such as soil erosion and pollution, encroachment on agricultural and ecological land, and traffic congestion (Yeh and Li, 1999; Seto et al., 2002). According to the official data, the urbanized area of this city was only 99 km^2 in 1988, but increased to 684 km^2 in 2001 and 1211 km^2 in 2010. A further worsening problem of land resources is that a large percent of remaining land belongs to sensitive or unavailable areas, such as water, hilly areas, and important agricultural or ecological land. The diminishing land stock raises a major question on how to reserve enough land for satisfying future social, economic and ecological demands. In 2009, Dongguan promulgated the ordinance of Eco-designated Line of Control (ELC) (Act 112) for ecological land protection. This ordinance requires a total of 1103 km^2, or about 44.7% of its total land area, to be strictly protected for ecological uses. However, the urban land use already claimed 49.1% of the total land area in 2010. The implementation of ECL will nevertheless face with severe challenges because of continuing economic growth and urban expansion.

3.2. Building the CA model

The 2006, 2010 TM images, which are with a spatial resolution of 30 m, were used to provide the empirical information about land use dynamics (the dependent variable) for calibrating CA. Radiometric and geometric corrections were carried out before classifying these images. The classification was based on a series of techniques, such as object-based classification, manual editing, and intensive field labeling with GPS. The image segmentation produced objects by aggregating similar pixels. Samples (objects) were then manually collected for each land use category (e.g. urban area, farm land, forest, water, fishpond, and bare soil). The average accuracies of the classification for these images are about 83–85% according to our field checking (Chen et al., 2011).

Land use changes are partially dependent on a series of spatial variables (the independent variables), such as various distances to attraction centers. These variables are often used to estimate the probabilities of land use changes during urban and land use simulation (Wu and Webster, 1998; Li et al., 2008). In this study, a number of spatial variables are defined as follows: 1) distance to the city center (*DisCity*), 2) distance to the town centers (*DisTown*), 3) distance to the railways (*DisRail*), 4) distance to the expressways (*DisExpress*), and 5) distance to the roads (*DisRoad*). It is rather

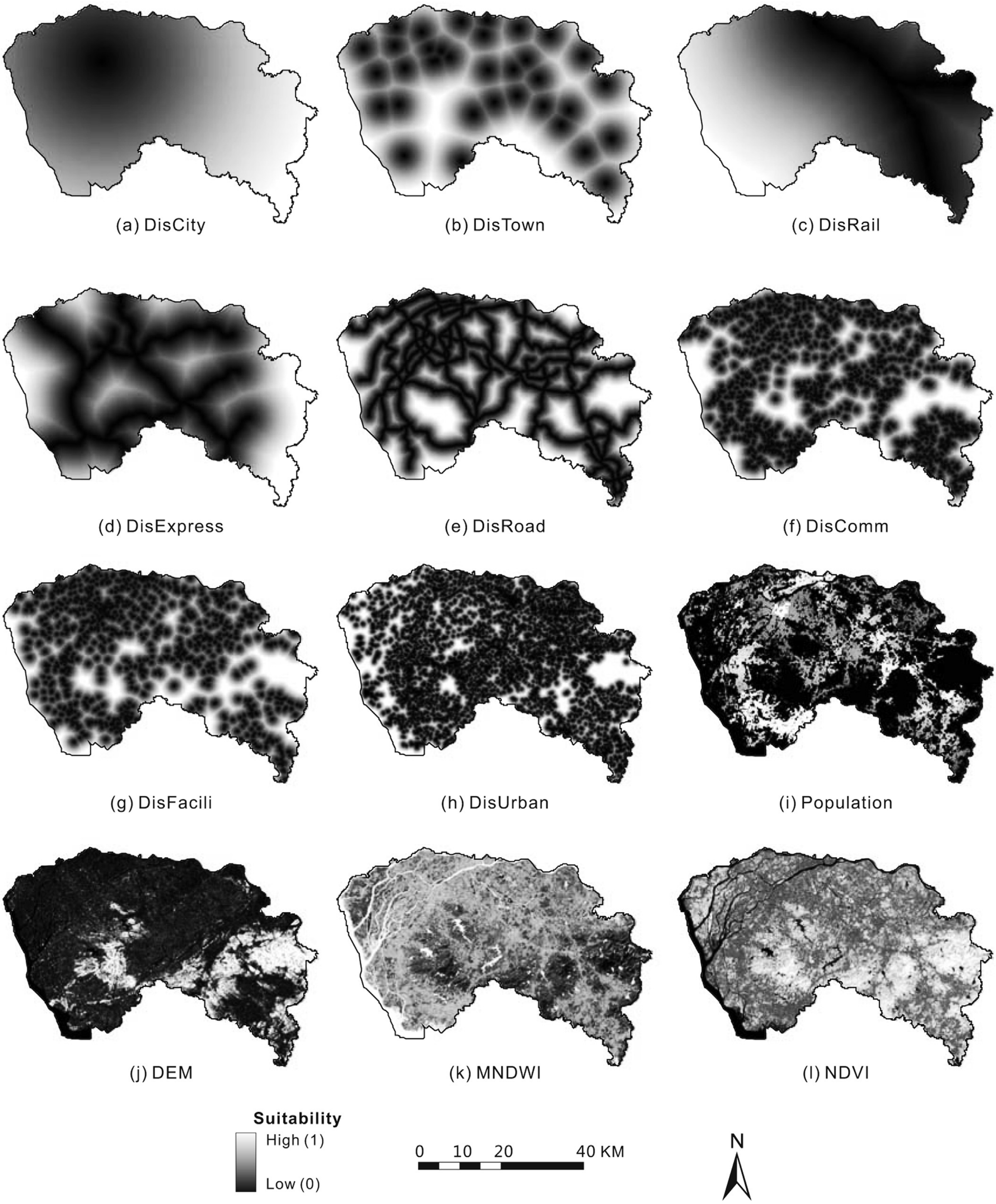

Fig. 2. Various spatial variables for CA and ANN models: (a) distance to the city center (DisCity), (b) distance to the town centers (DisTown), (c) distance to the railways (DisRail), (d) distance to the expressways (DisExpress), (e) distance to the roads (DisRoad), (f) distance to the business centers (DisComm), (g) Distance to facilities (DisFacili), (h) Distance to Urban Areas (DisUrban), (i) Population, (j) DEM, (k) MNDWI, and (l) NDVI.

Table 1
Accuracy of the logistic regression for the training and testing data.

	Converted	Non-converted	Accuracy
1. Training data			
Converted	17,905	7095	71.62%
Non-converted	11,347	13,653	54.61%
Total	29,252	20,748	63.12%
2. Testing data			
Converted	17,733	7267	70.93%
Non-converted	11,497	13,503	54.01%
Total	29,230	20,770	62.47%

Table 2
Accuracy of the neural network for the training and testing data.

	Converted	Non-converted	Accuracy
3. Training data			
Converted	4228	772	84.56%
Non-converted	456	4554	90.90%
Total	4684	5326	87.72%
4. Testing data			
Converted	4209	791	84.18%
Non-converted	554	4446	88.92%
Total	4763	5237	86.55%

convenient to create these variables by using common GIS (Geographical information systems) functions (Fig. 2).

The logistic-CA was built according to these dependent and independent variables. Logistic regression was implemented by using the weka-3-6-6 software (Hall et al., 2009). The overlay of classified 2006 and 2010 TM images reveals that there are 374,811 pixels and 1,458,528 pixels of converted (urbanized) and non-converted land respectively. For each land use type, 50,000 pixels were randomly drawn for the logistic regression (Wu, 2002; Li et al., 2008). The features for the regression include the labeled land use types (the dependent variable) and the above spatial variables (the independent variables). These samples were equally divided into two groups, the training data and the testing data. The total accuracies are 63.12% and 62.47% for the training data and the testing data respectively (Table 1). This regression yielded the parameters of the combined variable (z_{ij}^t) as described in Equation (2):

$$z_{ij}^t = 1.402 - 0.520x_{\text{CityProper}} - 0.840x_{\text{TownCentre}} - 0.232x_{\text{Railways}} - 0.718x_{\text{Expressways}} - 3.033x_{\text{Roads}} \quad (6)$$

where $x_{\text{CityCentre}}$, $x_{\text{TownCentre}}$, x_{Railways}, $x_{\text{Expressways}}$, and x_{Roads} represent the distance to the city center, the distance to the town centers, the distance to the railways, the distance to the expressways, and the distance to the roads respectively.

3.3. Building the ANN model

The SPOT images in 2010 were classified to provide the empirical information of illegal development for training the ANN model. Classified illegal development was verified by intensive labor work, such as field checking with differential GPS. Although much more labor costs are required to classify these SPOT images, it is worthwhile to use these high-resolution images for producing satisfactory classification of land use types. Illegal development was obtained by the overlay of these classified images with the planning maps from the Planning Department of Dongguan.

Illegal development is usually in leap-frog or fragmented patterns, characterized by nonlinear and complex behaviors. These patterns are more complex than those of general urban development. Therefore, more spatial variables should be incorporated in the prediction model to capture such highly nonlinear features. Two groups of spatial variables, proximity variables and site properties, are used as the inputs to this ANN model. Selection of these variables is based on previous studies (Verburg et al., 2006; White et al., 1997; Li and Yeh, 2002; Wu, 2002). The first group of variables (proximity variables) is the same as those used for CA (Li and Yeh, 2002). The second group of variables includes some additional proximity variables and site properties. These variables are: 1) distance to the business centers (*DisComm*), 2) distance to facilities (*DisFacili*), 3) distance to urban areas (*DisUrban*), 3) population, 4) *DEM*, 5) *MNDWI*, and 6) *NDVI* (Fig. 2).

The site properties were obtained by using remote sensing or GIS data. For example, the Normalized Difference Vegetation Index (*NDVI*) which allows for better identification of non-urbanized (developed) pixels is calculated according to the following equation (Tucker, 1979):

$$NDVI = \frac{\text{TM4} - \text{TM3}}{\text{TM4} + \text{TM3}} \quad (7)$$

where TM 3 and TM4 are the band 3 and 4 of Landsat TM data respectively.

The modified Normalized Difference Water Index (*MNDWI*) is used to identify the water pixels which should be excluded from development sites (Li et al., 2011). *MNDWI* is calculated as follows (Xu, 2006):

$$MNDWI = \frac{\text{TM2} - \text{TM5}}{\text{TM2} + \text{TM5}} \quad (8)$$

where TM 2 and TM5 are the band 2 and 5 of Landsat TM data respectively.

There are 19,513 pixels and 1,019,073 pixels of converted (illegal land) and non-converted land respectively. For each land use type, 10,000 pixels were randomly drawn for building the ANN model. These samples were equally divided into two groups, the training data and the testing data. The BP neural network was adopted by using the weka-3-6-6 software (Hall et al., 2009). The neural network consists of 12 neurons (representing the independent variables mentioned before) in the input layers, 7 neurons in the hidden layer, 2 neurons (representing the probabilities of converted or non-converted for the prediction) in the output layers. The number of the neurons in the hidden layer is decided by the default setting of the weka-3-6-6 software. This number is equal to the average of the number of input neurons and the number of output neurons ((12 + 2)/2 = 7). The weights of ANN are automatically determined according to the backpropagation algorithm which is to minimize the prediction error. Studies indicate that such "black box" approach has no physical meanings of internal parameters and physical relations between the parameters and output (Li and Gu, 2003). In our experiments, the total accuracies of ANN are 87.72% and 86.55% respectively by using the training data and testing data for predicting the illegal development (Table 2).

3.4. Early warning of illegal development based on a multi-model approach

These two models will not agree well with each other because of using different model structures and training schemes. If integrated properly, however, they can be complementary for improving the performance of the early warning. A convenient way to combine them is based on the simple "union" ("or") operation. The integration requires that these models are constrained by the same amount of land consumption for future years.

Table 3
Examples of ground truth data of illegal development measured by the high-precision GPS.

Patch no.	Location of the centroid	Land use	Area ($10^3 \times m^2$)
1348	113°45′28.02″E 22°56′07.57″N	Residential use	0.616
1367	113°43′01.88″E 22°58′12.44″N	Residential use	2.870
1369	113°40′42.42″E 22°58′32.60″N	Commercial use	1.154
1377	113°45′03.07″E 22°56′11.86″N	Commercial use	30.796
1392	113°40′50.63″E 22°58′42.04″N	Industrial use	4.414
1394	113°41′45.02″E 22°58′51.00″N	Commercial use	13.643
1407	113°41′04.47″E 22°59′02.76″N	Industrial use	7.385

First, the proposed CA was used to simulate the land development in 2015, 2020 and 2025 respectively. The predicted illegal development sites were identified by overlaying the simulated results with the Eco-designated Line of Control (ELC) and the Planning Red Line (PRL, or authorized blueprints). Those development sites are considered as illegal if they are within ECL and outside PRL according to the definitions of planning departments.

Then ANN was also used to provide alternative prediction of illegal development for future years. As illegal development sites are usually in highly non-linear or fragmented patterns, more spatial variables should be included as the explanatory variables to improve the predictability of EWS. In this ANN model, additional spatial variables are used in the input neurons for improving the prediction of illegal development. As the output layer only yields the probability of illegal development, a threshold value is used to determine if a site will be converted as illegal development. In this study, this threshold is decided in such a way that the total amount of illegal development by ANN will be equal to the predicted by CA for a future year. It is because ANN which is not a process model should be linked to CA for obtaining such process information.

In this study, the above CA, ANN and CA + ANN models were implemented based on the spatial resolution of 30 m because of using TM data. The ground truth data were acquired by using high-precision GPS. The GPS based on the Continuous Operational Reference System (CORS) provides up to centimeter-level accuracy for validating the warning results. Table 3 shows the examples of ground truth data of illegal development measured by the high-precision GPS. Each site of illegal development records the geometric and attribute information in terms of its location, size, and land use type. For example, Fig. 3 displays two examples of warning sites which are confirmed as illegal development by the field investigation. They are illegal residential plots and manufactory buildings which have not been approved by the planning department. The total area of these two illegal sites is 6735.2 m^2 and 7384.5 m^2 respectively, recorded by the high-precision GPS. The incorporation of GPS allows accurate geometric properties of illegal development to be delineated on the ground, and thus provides important information for carrying out legal actions against these activities. The accuracies of CA, ANN and CA + ANN (the integrated model) for the warning of illegal development were estimated by using the 2010 ground truth data (Table 4).

Figs. 4–6 show the predicted warnings of illegal development in a zoom-in area for 2010 by using CA, ANN and CA + ANN models respectively. These figures indicate that CA + ANN can effectively reduce the possibility of missing warnings. Fig. 4 clearly demonstrates that the CA model alone will produce a lot of missing

Fig. 3. Two examples of warning sites which are confirmed as illegal development according to the CORS GPS.

Table 4
Accuracies of CA, ANN and CA + ANN for the warning of illegal development based on the ground true data in 2010.

	Confirmed (true)	False	Missing	*FAR*	*MAR*
	In hectare				
CA	327.7	5234.4	1428.5	94.11%	81.34%
ANN	1376.46	5884.47	366.48	81.04%	21.03%
CA + ANN	1466.55	9555.84	276.39	86.69%	15.86%

$FAR = W_{false}/(W_{true} + W_{false})$
$MAR = W_{missing}/(W_{true} + W_{missing})$

warnings (blue color in the figure) (in web version). Fig. 5 indicates that the ANN model has better results than the CA for reducing the amount of missing warnings. However, the combined CA + ANN model can have better performances of reducing missing warnings than the single CA and ANN models (Fig. 6).

The proposed integrated model seems to perform much poorly with regard to the value of *FAR*. Actually, all of these models may be quite disappointed in terms of *FAR* because their *FAR* values are quite high (Table 4). As mentioned before, the use of *FAR* may result in missing warnings which can cause severe ecological and environmental risks or catastrophe in most situations. The analysis indicates that the proposed model performs quite well in terms of *MAR*. Compared with CA and ANN, this integrated model can increase the value of *MAR* by 65.48% and 5.17% respectively (Table 4). Therefore, this integrated model is more effective for warning potential illegal development at an early stage.

After the calibration, the proposed model was used to predict the warnings of illegal development for the years of 2015, 2020 and 2025 respectively. Fig. 7 shows the predicted warnings of illegal development in 2015 for Dongguan by using CA + ANN. It is interesting to find that illegal development is usually located near the edges of ECL (Fig. 8). For example, 86.9% of the illegal development in 2010 was identified in the buffer of 600 m from the edge of ECL. Stronger law enforcement actions should be conducted to protect these areas from potential development based on the warning results.

4. Conclusion

Rapid urban development has caused the significant loss of agricultural and ecological land in many cities in China. In 2005, Shenzhen implemented the first ordinance of Eco-designated Line of Control (ELC) which is known as "wired" ecological control line. The study area, Dongguan, is the second city to implement such policy to protect its shrinking land resources. According to government reports, however, such policy is faced with severe challenges because of proliferated illegal development in these fast growing regions. The successful implementation of ELC requires the strict prohibition of land development within eco-designated areas, supported by intensive monitoring efforts.

It is useful to develop an early warning system (EWS) to predict the threats to the regional ecological security by illegal development. A single model may not perform well for predicting illegal development which is in highly non-linear or fragmented patterns.

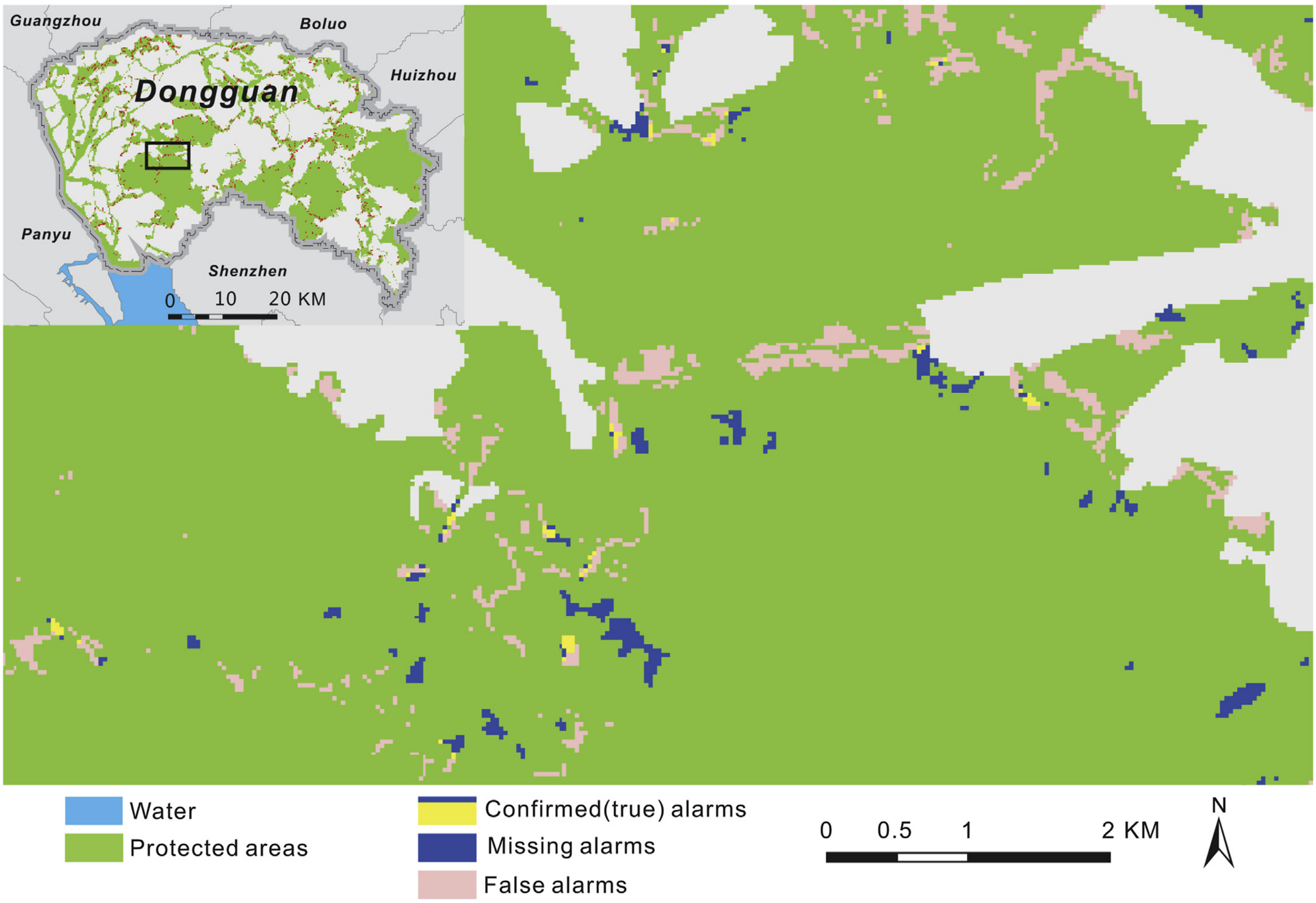

Fig. 4. Warnings of illegal development in 2010 for a zoom-in area by using CA.

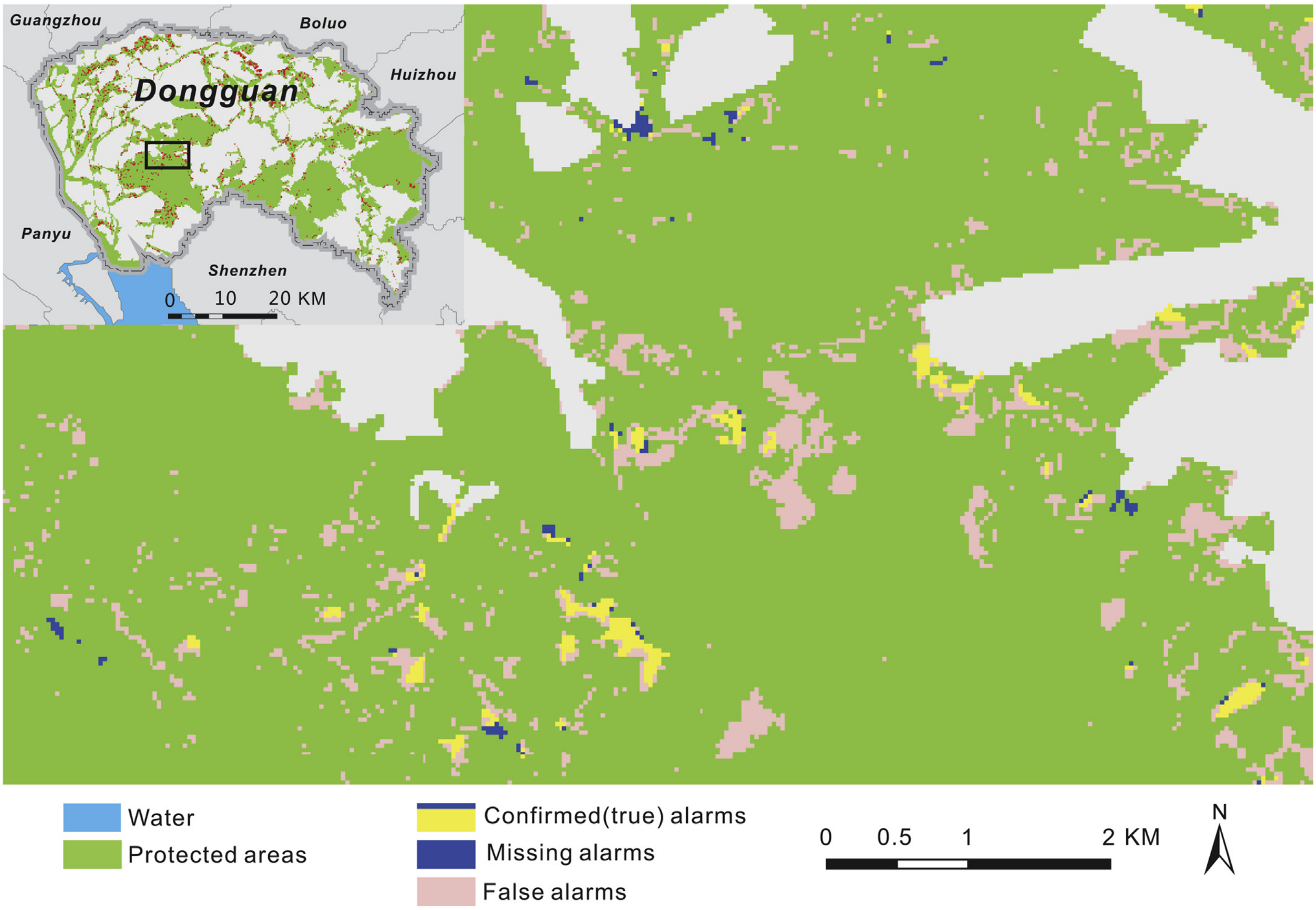

Fig. 5. Warnings of illegal development in 2010 for a zoom-in area by using ANN.

This study has demonstrated that a multi-model approach can help to increase the accuracy for the early warning of illegal development. In this study, cellular automata (CA) and artificial neural networks (ANN) are integrated to estimate the risks posed to natural area protection by rapid urban development.

The warnings of illegal development are obtained more accurately by combined use of CA and ANN models. Any illegal development sites predicted from either of these two models for future years will be considered as the warnings. Empirical data about illegal development are obtained from temporal remote sensing data, such as TM and SPOT satellite images. The GPS which is based on Continuous Operational Reference System (CORS) provides a powerful tool to verify the warning results with detailed and accurate geometric information. The warnings from the proposed model are used to identify the target areas for detailed field investigation. This significantly reduces labor costs and allows the ground measurement with high-precision GPS is feasible.

Equipped with the simulation models (CA and ANN) and the GPS tool, the proposed EWS is effective for understanding the process of illegal development at site-specific or sub-regional scales. The development of EWS can ensure better ecological control and allow for the evaluation of ecological viability and benefits. By issuing warnings of illegal development in advance, government officers can assign more monitoring actions in the potential places to prevent such infringement. This integrated system may help to find out alternative management scenarios, assess their impacts, and monitor the plan which is at an implemented phase. This EWS is also useful for identifying the driving forces which are responsible for illegal development, and the sequences of various development strategies.

In this study, the traditional concept about accuracy may not be applied to this proposed system. It is because the ultimate goal is not just to predict illegal development accurately, but rather to implement protection measures and legislation in the target areas. This needs to identify the potential sites which may have a chance to develop illegally. Compared with missing alarms, over-alarms or false-alarms of illegal development may not be a serious problem since ground checking can be carried out to verify or correct the prediction errors if they are not too much.

The experiments have indicated that the proposed integrated model can effectively reduce the possibility of missing warnings. Missing alarms should be a major concern of the early warning because uncaught sites of illegal development could cause severe ecological risks or damages. *MAR* (the ratio of missing alarms to all the warnings) is a much better indicator for the assessment of the performance of EWS than *FAR* (the ratio of false alarms, or unverified warnings to all the warnings). Compared with CA and ANN, this proposed model can increase the value of *MAR* by 65.48% and 5.17% respectively. Therefore, it is worthwhile to combine CA and ANN for producing satisfactory results of early warning.

The proposed method will be in favor for a certain amount of over-alarms or false-alarms. It is obvious that intensive labor work is required for the verification of alarms. The future studies need to find out what is the acceptable amount of over-alarms or false-

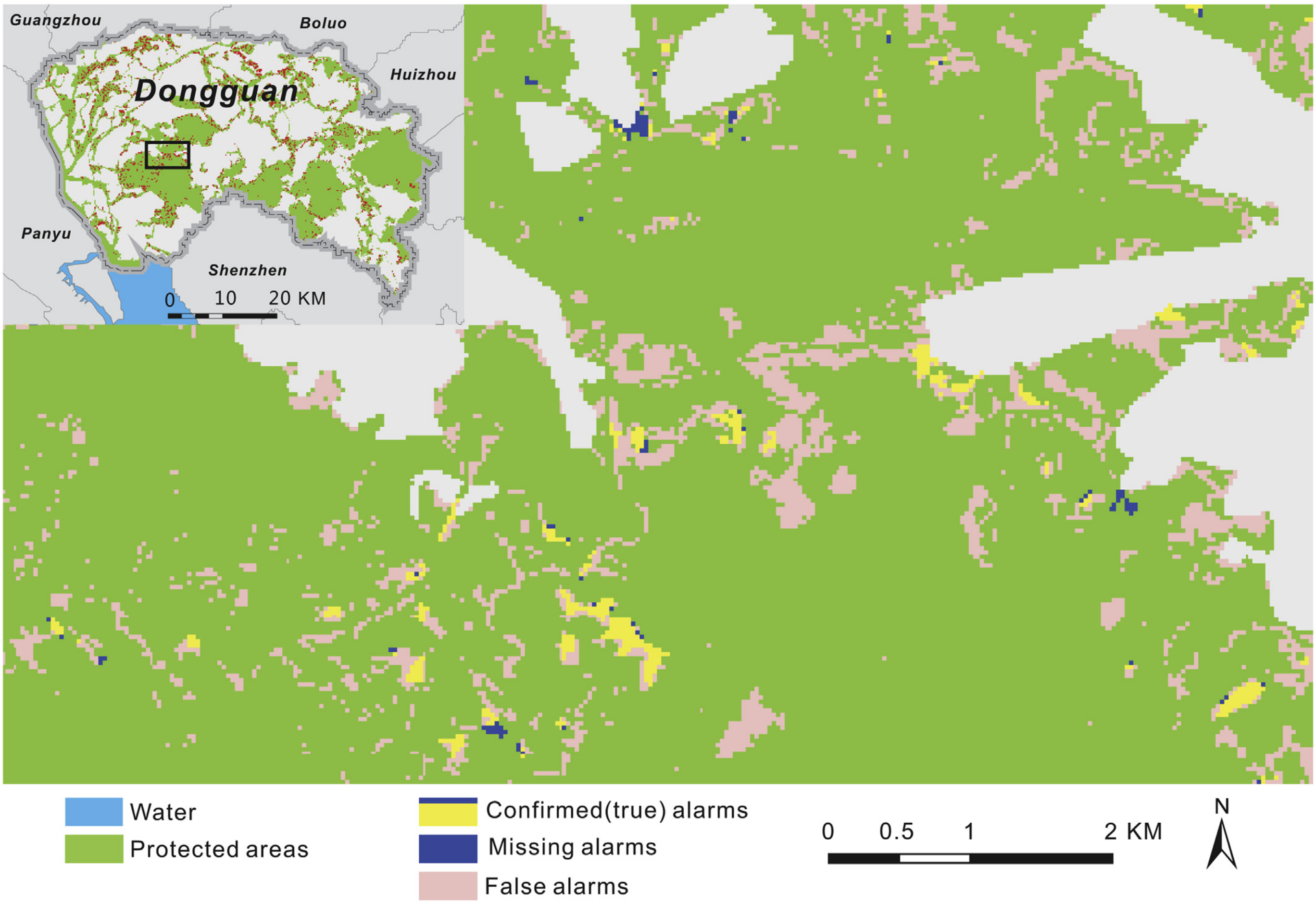

Fig. 6. Warnings of illegal development in 2010 for a zoom-in area by using CA + ANN.

Fig. 7. Warnings of illegal development in 2015 for Dongguan by using CA + ANN.

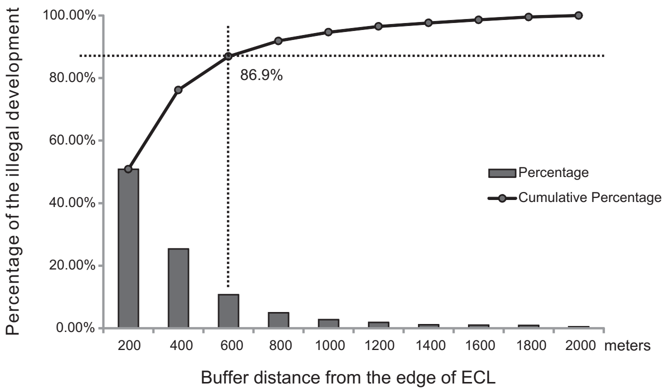

Fig. 8. The relationships between the illegal development in 2010 and the buffer distance from the edge of ECL.

alarms with regard to labor costs. Model parameters or thresholds can then be defined for improving the efficiency of EWS. Moreover, existing CA and other land use models are basically stationary because of using fixed transition rules. Studies have shown that such assumption can capture complex nonlinear dynamic behavior (Couclelis, 1988; Batty and Xie, 1994; White et al., 1997). However, future studies should focus on the use of updated data or assimilating techniques to improve the prediction capability of these models for far future years. It is also useful to see if the proposed approach can be generalized for the use in other cites/regions.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grant No. 41371376).

References

Barlindhaug, S., Holm-Olsen, I.M., Tommervik, H., 2007. Monitoring archaeological sites in a changing landscape-using multitemporal satellite remote sensing as an 'early warning' method for detecting regrowth processes. Archaeol. Prospect. 14, 231–244.
Batty, M., Xie, Y., 1994. From cells to cities. Environ. Plann. B 21, 31.
Barnes, L.R., Gruntfest, E.C., Hayden, M.H., Schultz, D.M., Benight, C., 2007. False alarms and close calls: a conceptual model of warning accuracy. Weather Forecast. 22, 1140–1147.
Chen, S., Billings, S.A., 1992. Neural networks for nonlinear dynamic system modelling and identification. Int. J. Control 56, 319–346.
Chen, Y., Li, X., Zheng, Y., Guan, Y., Liu, X., 2011. Estimating the relationship between urban forms and energy consumption: a case study in the Pearl River Delta, 2005–2008. Landsc. Urban Plann. 102, 33–42.
Clarke, K., Hoppen, S., Gaydos, L., 1997. A self-modifying cellular automaton model of historical. Environ. Plann. B: Plann. Design 24, 247–261.
Conway, T.M., Wellen, C.C., 2011. Not developed yet? Alternative ways to include locations without changes in land use change models. Int. J. Geograph. Info. Sci. 25, 1613–1631.
Couclelis, H., 1988. Of mice and men: what rodent populations can teach us about complex spatial dynamics. Environ. Plann. A 20 (1), 99–109.
Deng, X., Huang, J., Rozelle, S., Uchida, E., 2008. Growth, population and industrialization, and urban land expansion of China. J. Urban Econom. 63, 96–115.
Dompka, V., 1996. Human Population, Biodiversity and Protected Areas: Science and Policy Issues. American Association for the Advancement of Science, Washington, DC.
Ezeonu, I.C., Ezeonu, F.C., 2000. The environment and global security. The Environmentalist 20, 41–48.
Foody, G.M., 1996. Relating the land-cover composition of mixed pixels to artificial neural network classification output. Photogramm. Eng. Remote Sens. 62, 491–498.
Forman, R.T.T., 2008. The urban region: natural systems in our place, our nourishment, our home range, our future. Landsc. Ecol. 23, 251–253.
Gong, J., Liu, Y., Xia, B., Zhao, G., 2009. Urban ecological security assessment and forecasting, based on a cellular automata model: a case study of Guangzhou, China. Ecol. Model. 220, 3612–3620.
Grossberg, S., 1988. Nonlinear neural networks: principles, mechanisms, and architectures. Neural. Netw. 1, 17–61.
Hockey, P.A.R., Curtis, O.E., 2009. Use of basic biological information for rapid prediction of the response of species to habitat loss. Conserv. Biol. 23, 64–71.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. 11 (1), 10–18.
Li, B., 2007. Loss of 10 Percent of Ecological Protected Land [online] <http://news.qq.com/a/20070214/000576.htm> (Chinese).
Li, X., Chen, Y., Liu, X., Li, D., He, J., 2011. Concepts, methodologies, and tools of an integrated Geographical Simulation and Optimization System (GeoSOS). Int. J. Geograph. Info. Sci. 25, 633–655.
Li, X., Yang, Q., Liu, X., 2008. Discovering and evaluating urban signatures for simulating compact development using cellular automata. Landsc. Urban Plann. 86, 177–186.
Li, X., Yeh, A.G.O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. Int. J. Geograph. Info. Sci. 16, 323–343.
Li, Y., Sun, X., Zhu, X., Cao, H., 2010. An early warning method of landscape ecological security in rapid urbanizing coastal areas and its application in Xiamen, China. Ecol. Model. 221, 2251–2260.
Li, Y.T., Gu, R.R., 2003. Modeling flow and sediment transport in a river system using an artificial neural network. Environ. Manage. 31 (1), 122–134.
Lin, Y.P., Chu, H.J., Wu, C.F., Verburg, P.H., 2011. Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling – a case study. Int. J. Geograph. Info. Sci. 25, 65–87.
Liu, J., Linderman, M., Ouyang, Z., An, L., Yang, J., Zhang, H., 2001. Ecological degradation in protected areas: the case of Wolong Nature Reserve for giant pandas. Science 292, 98–101.
McDonald, A.A., Liu, J., Prince, H., Kress, K., 2001. A socio-economic-ecological simulation model of land acquisition to expand a national wildlife refuge. Ecol. Model. 140, 99–110.
Openshaw, S., 1998. Neural network, genetic, and fuzzy logic models of spatial interaction. Environ. Plann. A 30, 1857–1872.
Quansah, J.E., Engel, B., Rochon, G.L., 2010. Early warning systems: a review. J. Terrestr. Obs. 2, 24–44.
Seto, K.C., Woodcock, C., Song, C., Huang, X., Lu, J., Kaufmann, R., 2002. Monitoring land-use change in the Pearl River Delta using Landsat TM. Int J Remote Sens 23, 1985–2004.
Tang, W.S., Chung, H., 2002. Rural-urban transition in China: illegal land use and construction. Asia Pac. Viewp. 43, 43–62.
Tegler, B., Sharp, M., Johnson, M.A., 2001. Ecological monitoring and assessment network's proposed core monitoring variables: an early warning of environmental change. Environ. Monit. Assess. 67, 29–55.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.
Verburg, P.H., Overmars, K.P., Huigen, M.G.A., de Groot, W.T., Veldkamp, A., 2006. Analysis of the effects of land use change on protected areas in the Philippines. Appl. Geo. 26, 153–173.
Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., Mastura, S.S.A., 2002. Modeling the spatial dynamics of regional land use: the CLUE-S model. Environ. Manage. 30, 391–405.
Wang, D.S., Li, K.H., 2008. Illegal land development from local governments. Public Adm. 8, 21–23.
White, R., Engelen, G., Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. Environ. Plann. B 24, 323–344.
Wong, S.W., Tang, B., 2005. Challenges to the sustainability of 'development zones': a case study of Guangzhou Development District, China. Cities 22, 303–316.
Wu, F., 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. Int. J. Geograph. Info. Sci. 16, 795–818.
Wu, F., Webster, C.J., 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environ. Plann. B 25, 103–126.
Wu, F., Yeh, A.G.O., 1999. Urban spatial structure in a transitional economy. J. Am. Plann. Assoc. 65, 377–394.
Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033.
Yeh, A.G.O., Li, X., 1999. Economic development and agricultural land loss in the Pearl River Delta, China. Habitat Int. 23, 373–390.
Zhao, P., 2011. Managing urban growth in a transforming China: evidence from Beijing. Land Use Pol. 28, 96–109.
Zhao, P., Lü, B., Woltjer, J., 2009. Consequences of governance restructuring for quality of urban living in the transformation era in Beijing: a view of job accessibility. Habitat Int. 33, 436–444.