Study on the driving forces and prediction of built-up area for urban expansion in Kunming

Yulian Hong, Jianhua Xu
The Research Center for East-West Cooperation in China,
East China Normal University
Shanghai, China
hongrema@163.com jhxu@geo.ecnu.edu.cn

Zhanyong Wang
Department of Spatial Information Science and Engineering,
Xiamen University of Technology
Xiamen, China
wangzhy@xmut.edu.cn

Abstract—**Based on the analysis of driving forces of urban land expansion by Principal component analysis (PCA), this paper established a predicting model of urban built-up area for future by using socio-economical data. Being good at the performance of nonlinear approximation, artificial neural network (ANN), especially the back propagation algorithm (BP), is applied in the prediction of bulit-up land and had attained satisfactory results. Taking Kunming for example, the results showed that the urbanization is the decisive factor influencing urban land expansion, and a predicting model combined PCA and BP-ANN used to predict urban built-up area in the year of 2009—2015. The method employed in this paper can provide a reference to study on urban land expansion for urban development and planning in the inland cities lacking of multi-sources data.**

Keywords- urban land expansion; driving forces; PCA; BP Neural Network; built-up area; Kunming

I. INTRODUCTION

In the past several decades, land use/land cover change has been a key subject for the study of global change or global warming. Urban land expansion is one of the most directly representative forms of land use/land cover change, especially in the change of land use pattern and urban space distribution because of land, social and economic pressure [1]. It is also known as that city is the center of regional system acting as a highly influential complex system on spatial and temporal, and urban land expansion has been one of the key subjects for study on land use change. At the same time, urban land use is the supporter for socio-economy development of city and the scale is bound to increase with the fast development of urbanization. On the other hand, urban land expansion is the necessary requirement for the economy development and the improvement of nation livelihood [2]. So an accurate estimate of the demand of urban land area is the foundation of decision-making for city plan and land using planning, which can coordinate the other developmental needs of the nation.

For urban land expansion, this paper mainly considered the expansion of built-up area, because it is easy to calculate and convenient to count [3]. Based on socio-economic data, this paper has analysed the driving forces of the expansion of built-up area in Kunming by principal component analysis (PCA) quantitatively and predicted the built-up area in the period of "Twelfth Five-year Plan" combined with ACP and BP-ANN, which can provide references to urban land planners. Additionally, the results would be meaningful to study the city extension processes and be helpfull to study on urbanization in inland cities.

II. STUDY AREA AND METHODOLOGY

A. Study Area

Kunming(102°10´~103°41´E, 24°24´~26°33´N), an inland city, is located in the center of Yunnan plateau, Southwest China, taking an elevation between 1500-2800m. It belongs to the highland landform, which takes up 87.5% of the total territory, but the roof of the plateau is relatively flat. The topography mainly inclines from north to south and the mid-land is obviously higher than both east and west sides. Kunming faces with a lake on one side and is embraced by mountains on the other three sides. This city situates in the low latitude mountain plateau monsoon climate, with long sunlight, short frost season, and an average temperature of 15°C, so it is famous for "Spring City". The excellent geographic location and climate advantages enable Kunming to be the political, economic & cultural center and transportation hub in Yunnan province. Meanwhile, Kunming acts as the southwest portal of China aiming to be a regional international city to Southeast and South Asia.

History statistics showed that the built-up urban area of Kunming is increased with fast economic development and deepening urbanization, simultaneously, there were numerous agricultural lands to shift into non-agricultural lands and the urban dimension was drastically scaling out in the form of development areas and new towns. Accurately, built-up area of Kunming is about 32 km^2 at the middle of the 1980s, but to 276 km^2 in 2008, which shows an increase by 8.63 times of that in 1986, and an annual change rate of 10.52% in the past 23 years. Compared to the other inland cities in China, the rapid urban land expansion of Kunming has both particularity and representativeness, fully reflecting the similarities of the driving forces among these cities.

978-1-4244-9171-1/11/$26.00 ©2011 IEEE

B. Data and method

This paper mainly has picked 31 socio-economic indexes of 23 sample data (1986-2008) from Kunming Statistical Yearbook to explain the expansion of urban land. The indexes contain the population, social structure, economic developing, consumption structure, indirect indexes of policy influence and transportation, so the system is dominative and comprehensive, which can weaken the influence of subjective thought by quantitative analysis [4].

Based on the representative factors, firstly we predict the values of them in the next seven years (2009-2015) using grey prediction, multiple regression, exponential smoothing and so on [5]. Then, we extract a few independent and integrated principal components by PCA. Finally, based on the data which is reduced dimensions, a BPNN model is established to predict the built-up area demand in the next several years.

1) Principal component analysis(PCA)

PCA is a mathematic method to eliminate dimensions, which provides a simple way to couple with cross data. Actually, PCA extracts a smaller set of underlying new variables that are uncorrelated, mutually independent and mathematically represented by linear combinations of original variables in the X matrix [6]. These new variables, taken as principal components (PCs), are calculated to account for most of the variance exsiting in the X matrix. PCA decomposes the data matrix X as presented in (1).

$$X = \sum_{i=1}^{k} t_i \cdot p_i + E \quad (1)$$

where k is the number of samples in the X data set, and t_i vectors are known as scores (i.e. values) on the PCs (i.e. new variables) extracted by PCA. The p_i vectors are known as loadings and contain information on how the variables relate to each other. The scores t_i can be interpreted as projections of the variable to a new space spanned by the PCs [6] (i.e. when the original variables are transformed to PCs, each variable in the X matrix is projected on to the PC space). Generally, the first few principal components can explain most of the variability of the original data.

In order to characterize the driving force of Kunming's urban land expansion, three PCs, i.e. the first, second and third PCs were extracted from 31 driving factors in this paper. Then, based on the load values, we can get the driving factors and analyze the main driving force of urban land expansion.

2) Back Propagation Artificial Neural Network (BP-ANN) model

BP-ANN, a representative and prevalent model among ANNs, is a complex neural network which comprises a large set of simple neural cells, and it is a multi-layer and feed-forward network with the network training based on error back-propagation and weight or threshold adjustment [7]. In BP-ANN, "learning" is a supervised process that occurs with each cycle of "epoch" (i.e. each time the network is presented with a new input pattern) through a forward activation flow of inputs and the backwards error propagation of weight adjustment.

In this study, we adopt a fully conjunctive three-layer network with input layer, hidden layer and output layer to predict built-up area according to the following principles. It has been proved theoretically that a three layers network can approximate to any given continuous functions.

Suppose $X_k=(x_1,x_2,\ldots,x_n)$, $k=1,2,\ldots,m$, is the sample vector for inputs with the training sample number m and the neuron number of input layer n, moreover, $Y_k=(y_1,y_2,\ldots,y_q)$, $k=1,2,\ldots,m$, is the desired output vector corresponding to the input sample with the neuron number q, so the inputs of nodes in hidden layer can be obtained according to (2) as follows:

$$S_j = \sum_{i=1}^{n} w_{ij} x_j - \theta_j \quad (2)$$

in which, w_{ij} is the weight linking i^{th} neuron in input layer to j^{th} neuron in hidden layer, θ_j is the threshold of j^{th} neuron in hidden layer.

The outputs of nodes in hidden layer can be obtained as (3), where p is the neuron number of hidden layer, and f is the excitation function expressed as (4) which can transform the input of a neuron in hidden layer or output layer into the output of this neuron with the purpose to simulate the nonlinear characteristics of biological neurons.

$$b_j=f(S_j) \; (j=1,2,\ldots,p) \quad (3) \quad / \quad f(x)=1/(1+e^{-x}) \quad (4)$$

While the inputs of nodes in output layer is calculated as (5):

$$L_t = \sum_{j=1}^{p} v_{jt} b_j - \gamma_t \quad (5)$$

where L_t is the inputs of t^{th} neuron in output layer, v_{ij} is the weight linking j^{th} neuron in hidden layer to t^{th} neuron in output layer, γ_t is the threshold of t^{th} neuron in output layer. Meanwhile, $C_t=f(L_t)$ is used to calculate the output of t^{th} neuron in output layer.

At the moment, an input mode has completed a positive sequence propagation process. In order to reduce the chance of falling into local minimum value and raise the speed of convergence of network model, an improved momentum BP algorithm is adopted to train network, and the training process can be described as (6),

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t+1) + \mu \Delta w_{ij}(t), \quad \Delta\omega_{ij}(t) = \eta \frac{\partial E}{\partial \omega} \quad (6)$$

where μ is the momentum factor, and E is the error function which can be expressed as (7) at t^{th} moment.

$$E(t) = \frac{1}{2}\sum_{j=1}^{q} (y_j(t) - d_j(t))^2 \quad (9)$$

Corresponding to j^{th} neuron, $y_j(t)$ is the actual output and $d_j(t)$ is the desired output. When $E(t) \le \varepsilon$ (ε is a given error in advance), BP network will stop training and the network model just is what we need.

III. RESULTS AND DISCUSSION

A. Driving force analysis of urban land expansion

Historical studies indicated that urban land expansion was the result of the joint effect of social, economic, and political factors [8]. An index system of driving factors was developed according to the previous discussion. In this paper, the driving force index system of urban land expansion consisted of 31 indexes [3][9]: total population (x_1, ten thousand), natural grow rate (x_2, %), employed person (x_3, ten thousand), total GDP (x_4, million yuan), per capita GDP (x_5, yuan), primary industrial outputs (x_6, million yuan), secondary industrial outputs (x_7, million yuan), tertiary industrial outputs (x_8, million yuan), local revenue (x_9, million yuan), real estate investment (x_{10}, million yuan), construction industry outputs (x_{11}, million yuan), per capita dweller wage (x_{12}, yuan), total outstanding of desposits (x13, million yuan), total retail sales of consumable (x_{14}, million yuan), total retail sales for state-owned economy (x_{15}, million yuan), total retail sales for collective (x_{16}, million yuan), total retail sales for individual (x_{17}, million yuan), total retail sales for catering (x_{18}, million yuan), professional health workers (x_{19}, thousand), Consumer Price Index (CPI, x_{20}), Retail Price Index (RPI, x_{21}), financial expenditure (x_{22}, million yuan), completion of infrastructure (x_{23}, million yuan), real estate development area (x_{24}, square meter), total customs import&export trade (x_{25}, million dollar), total travelling revenue (x_{26}, million yuan), total international tourism (x_{27}, ten thousand), total international travelling exchange revenue (x_{28}, million dollar), exchange of freight (x_{29}, ten thousand tons), exchange of passenger (x_{30}, ten thousand per kilometer), capacity of passenger carrying (x_{31}, ten thousand). Firstly, by using SPSS 13.0 software, correlation analysis was carried out (Table 1), and the results showed that all the selected indexes had significant relationship with the index of built-up area (significant level over 95%). Then, through PCA, the rotated component matrix with three principal components (PCs) was extracted, and the cumulative variance explained by the three principal components was 93.369%. PC1 had significant relationship (such as X1, X3-X14, X17-X18, X22-X25, X30-X31, variable loading over 0.8) with the indexes which could be summarized as urbanization. PC2 could be connected with the indexes of total retail sales for state-owned economy and collective (such as X15, X16 and X29, variable loading over 0.8), but the total retail sales for individual and catering were relatively low, so it might be as industrial structure. PC3 could be connected with the indexes of CPI and RPI (i.e. X20 and X21, variable loading over 0.8), which can be summarized as policy influence.

TABLE I. Correlation coefficient and load values of 31 indexes

Index	Correlation coefficient	Load value		
	Built-up area	*PC1*	*PC2*	*PC3*
X1	0.948**	0.924	0.257	0.178
X2	-0.591**	-0.526	-0.440	0.318
X3	0.981**	0.879	0.450	0.112
X4	0.987**	0.901	0.363	0.226
X5	0.988**	0.827	0.458	0.285
X6	0.986**	0.819	0.472	0.285
X7	0.985**	0.909	0.364	0.193
X8	0.987**	0.902	0.346	0.193
X9	0.793**	0.900	0.194	-0.094
X10	0.94**	0.953	0.244	0.137
X11	0.96**	0.940	0.275	0.188
X12	0.987**	0.909	0.336	0.229
X13	0.96**	0.953	0.239	0.180
X14	0.976**	0.929	0.309	0.193
X15	0.613**	0.284	0.807	0.132
X16	0.673**	0.276	0.838	0.330
X17	0.935**	0.974	0.155	0.151
X18	0.954**	0.915	0.220	0.288
X19	0.787**	0.622	0.635	0.014
X20	-0.442*	-0.211	-0.156	-0.925
X21	-0.482*	-0.184	-0.300	-0.891
X22	0.957**	0.938	0.287	0.150
X23	0.946**	0.952	0.263	0.116
X24	0.972**	0.918	0.332	0.184
X25	0.897**	0.947	0.260	-0.054
X26	0.944**	0.792	0.363	0.439
X27	0.924**	0.671	0.572	0.351
X28	0.926**	0.680	0.569	0.378
X29	0.591**	0.164	0.867	0.175
X30	0.984**	0.877	0.357	0.269
X31	0.939**	0.835	0.332	0.379

** and * correlation are significant at the 0.01 and 0.05 level (2-tailed), respectively.

Furthermore, we further investigated the socio-economic driving forces effecting Kunming's urban land expansion, and they could be demonstrated by three phenomena, aiming to assist the quantitative study on urbanization, and explain the actual urban expansion comprehensively.

1) Socio-economic development

Socio-economic development is the radical force of urban land expansion.The development of socio-economy increased residental income and purchasing power, leading to high demands of housing and continuous attraction of investment and labor force. A large number of floating population swarm into the city promoting the conversion of land use type and structure. On the other hand, pursuing for better living environment and with the fast development of real estate business, people migrate to suburbs accelerating the expansion. The development of tourist industry cannot be neglected to study Kunming's urban land expansion. Kunming is a worldwide well-known tourist city, and the characteristics of the expansion form are dominated by the utilization of landscape resources to a large extent.

2) Transportation influence

Improvement of traffic condition in each district has boosted to change the distribution of building construction impacting urban land expansion. The promotion of transportation system efficiency could also encourage the transformation of built-up area expanding from centralization to suburbanization, and enhance the business communication among districts. During the period of 1992-2000, Kunming's urban land expansion was developed both along the main traffic arteries and among them, in the 2000s, it is dominant by the horizontal filling. The allocation of town groups with close economic relationship also followed the traffic corridor [10].

3) Policy influence

Policy has predominant influence on urban land expansion through all stages, but it is a complex factor to represent. Obviously, the continuous extension of Kunming's districts presents the leaping growth of built-up area to some extent, e.g., Dongchuan city was evacuated, while Dongchuan District setted up in 1998. In the 2000s, Kunming has drawn up a series of construction plans of urban system due to the environmental problems, and proposed the strategy of "modern new Kunming", expanding the core area to southeastward and speeding up the space allocation and structure adjustment. In addition, the gradually loose control on household registration attracts numerous immigrants to Kunming.

B. BP network modeling and accuracy test

Based on the principle of BP-ANN model, this paper has constructed a three layer BP network model, taking socio-economic variables (X1-X31) as inputs while built-up area as output to train the neural network. In other words, the built-up area during the period of 1986-2006 used as the study sample set, and the multi-variables used as the simulation inputs, then train them year by year, furthermore, take the data of 2007 and 2008 as the test sample to examine the accuracy of the BP network model and compare the residual values between the fitting result and the raw data until the residual values of the two sequences are floated in a certain ranges. Last but not least, we can use the well trained BP network to predict the built-up area from 2009 to 2015.

The number of nodes for hidden layer is always determined by the optimal combination between initial learning time and precision. In this paper, based on trialing repeatedly, the structure of the network model is determined as 31-14-1, i.e. when use 3 hidden layer's neurons, and take Tansig and Purelin as input and output transfer function respectively, as well as apply the Learndm as the adaption learning function to prevent exceeding optimization, the test result showed that the built-up area of Kunming in 2007 and 2008 are 254.3864km^2 and 268.9537km^2 respectively. Through analyzing the error between test result and raw data, relative errors of the two years are -0.55% and 2.55%. Furthermore, in the simulated modle, the fitting average absolute error was almost bellow 1.23, while the relative error was less than 0.91 percent, which proved that using BP neural network to establish built-up area prediction model is practicable.

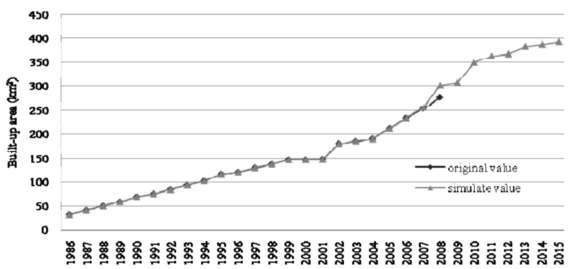

Figure 1. Simulation and prediction of built-up area in the period of 1986-2015

For BP-ANN, in the stages of model's learning and training, the network will adjust the each factor's weight automatically according to the different influences on the urban land expansion. Therefore, the application of BP-ANN can avoid the shortage for selecting the indexes [4].

IV. CONCLUSIONS

1) In the empirical study, the impact factors of built-up area can be summarized up as urbanization, industrial structure and policy influence. Especially, urbanization acts as the uppermost driving force. Then further analysis has been delivered in the development of socio-economy, transportation and policy influence to asist the development of quantitative study on urbanization comprehensively.

2) It's effective to comprehensively understand the driving forces of urban land expansion by PCA, which also plays a good role in noise filtering for the inputs of BP network model and can speed up the model convergence. Also, this research has proved that it can overcome the negative influence of coupling variables on convergence rate and model accuracy.

3) BP-ANN can well predict the built-up area in the term of nonlinear, and the prediction precision is much better than some traditional methods. Based on the three PCs, it has reduced the problem of error propagation and cumulative effectively, i.e. by decreasing the number of input neuron it can enhance the network performance and improve the prediction.

REFERENCES

[1] H. Alphan, H. Doygun, and Y. I. Unlukaplan, "Post-classification comparison of land cover using multitemporal Landsat and ASTER imagery: the case of Kahramanmara angstrom," Environmental Monitoring and Assessment. Turkey, vol. 151, pp. 327–336, 2009.

[2] B. Lei, "The comparison of multi-regression analysis model and BP neural network in predicting the urban construction land area," Urban studies, pp. 45–48, January 2008.

[3] Y. B. Liu, X. F. Song, "Study on the characteristics and driving forces of changes of the urban land-use structure in Wuhan in 1990s," Ecological Economy, vol. 1, pp. 57–60, Feburary 2005.

[4] Y. Y. Zhao, L. J. Pu and X. T. Hu, "Application of neural network in the prediction of urban construction land area", Resources and environment in the Yangtze Basin, vol. 15, pp. 14–18, January 2006.

[5] Z. Y. Wang, F. Lu and Y. Zhang, " Using the method combing PCA with BP Neural Network to predict water demand for urban development," IEEE International Conference on Natural Computation(ICIP-09). Tianjing, vol. 2, pp. 621–625, December 2009.

[6] S. Wold, K. Esbensen and P. Geladi, " Principal components analysis", Chemometrics and Intelligent Laborary Systems,2nd ed., pp. 37-52,1987.

[7] J. H. Xu, "Mathematical Methods in Contemporary Geography ," Higher Education Press, Beijing, 2002.

[8] P. H. Verburg and K. Kok, " A spatial explicit allocation procedure for modeling the pattem of land use change based upon actual land use," Ecological Modeling, vol. 116, pp. 45–61, 1999.

[9] L. Y. Shi, G. F. Shao, SH. H. Cui, and et al., "Urban three-dimensional expansion and its driving forces," Chinese Geographical Science, vol. 19, pp. 291–298, April 2009.

[10] W. B. Tan, B. Liu, and Z. X. Zhang, " Remote sensing monitoring and analyzing of the urban built-up land expansion in Kunming in 1974-2004," Geo-information science, vol. 11, pp. 117–124, January 2009.