Simulation of Oil Spill Using ANN and CA Models

Yihan Zhang[1], Jigang Qiao[1]*, Bingqi Wu[1], Weiqi Jiang[1], Xiaocong Xu[2], Guohua Hu[2]
[1]School of Geography and Tourism,
Guangdong University of Finance and Economics, Guangzhou, China
[2]Guangdong Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning,
Sun Yat-sen University, Guangzhou, China
*Corresponding author, e-mail: qjg821@263.net

Abstract—**In this paper, the artificial neural network (ANN) used to obtain transition rules in oil spill CA model. Model parameters are difficult to obtain in many traditional oil spill models, as they cannot meet the requirements of rapid response for oil spills. Therefore, a new oil spill model – ANN oil spill CA model was established in this paper. This model can simulate the change process of oil spill by setting initial image, verification image, and impact factors. Experimental results show that the simulation results have a good performance with overall accuracy of 96.6% and Kappa coefficient of 0.826. It was also found that the consistency of simulation results is proportional to the ratio of training sample. However, the higher the ratio of the training sample, the more computation is need in the ANN training. We also studied the effect of neurons number in the hidden layer. Studies show that the consistency of simulation results becomes better with the increase of neurons number in the initial stage for good fitting rate of training sample. However, the consistency of simulation results get worse for over-fitting of training sample in following stage.**

Keywords-cellular automata(CA), artificial neural network (ANN), simulation of oil spill, , DeepSpill

I. INTRODUCTION

Cellular automata (CA) models are a grid-based models with discrete space-time, finite state, and local rules. They are often used to simulate the evolution of self-organizing systems for strong computation ability [1]. A CA model has four elements, cells, states, neighborhood, and transition rules. The states determined by transition rules by using its previous state and neighborhood relations [2]. These transition rules are local, but they can be used to simulate complex and global geographical phenomena. This "bottom-up" concept is a good analysis tool for geographical evolution [3]. Moreover, CA models can couple with the data of remote sensing, geographic information systems (GIS). Therefore, CA models are widely used in geography, biology and other study fields [4-6].

CA models are also introduced into oil spill simulation. Oil is affected by many factors, including dynamic factors (wind, waves, and currents), environmental factors (temperature and salinity), characteristics of oil, and other factors. Oil spills in the ocean are a complex, dynamical process, which contains diffusion (physical expansion driven by gravity, surface tension, and inertia), efflorescence (chemical changes caused by evaporation, dissolution, emulsification, and biological degradation) and drift (dynamic movement caused by current and wind) [7-9]. Some scholars established numerical models based on physical processes of oil spill using differential equations. These models have important roles for studying mechanism and evolution of oil spill. However these models are complex and difficult to determine parameters. Some researchers tried to introduce cellular automata into simulation of oil spill. Karafyllidis firstly proposed the CA model to simulate oil spills by considering the effect of wind, current, and evaporation. Experiments showed that the simulation results are consistent with the real scenario of oil spills [10]. Rusinovic and Bogunović developed an improved version based on the CA model of Karafyllidis. Their model considered the effects of oil vertical transportation, evaporation, water and wind currents, dissolution, deposition, and emulsification. The model is used to simulate hypothetical oil slick behavior [11]. Wang et al defined the rules of oil spill and input into CA model. The model was applied in a water pollution accident. Drift trajectories are obtained successfully. However, the simulation of oil spill based on CA model is still in the initial stages. And most of the existing models are complex and difficult to obtain accurate model parameters.

In order to overcome the problem of obtaining accurate model parameters, the artificial neural network (ANN) used to obtain transition rules in oil spill CA model. Firstly, the ANN is trained by using training sample. The model parameters in ANN is defined after training, Then the trained ANN is used to simulate the change processes of oil spill. The model can avoid using the model parameters (diffusion coefficient, evaporation coefficient) which is difficult to obtain accurate value. The model is simple and it is very convenient to obtain accurate parameters of CA model. Besides, the model can simulate the complex oil movement by using a few inputs, such as the initial image, image of impact factors, and the weight of factors. In this paper, the proposed model was applied to simulate the surface behavior of oil spills in the DeepSpill project. Experiments showed that the presented model can obtain accurate dynamic changes of oil spill, which can meet the requirements of quick response in oil spill accident.

II. METHOD AND MATERIAL

A. Artificial neural network

Artificial neural network is a mature pattern recognition method, which simulates the behavior of the human brain. The information processing in ANN depends on the communication of neurons. The different respond (knowledge) is reflected by different neural network structures and weights [12]. The most common used training method in ANN is backward propagation

Supported by Science and Technology Program of Guangzhou, China (Grant No. 2013J410006), National Natural Science Foundation of China (Grant No. 41301408) and Guangdong Natural Science Foundation (Grant No. S2013040016071).

(BP) method. This kind of network contains an input layer, a hidden layer and output layer (Figure 1). Each layer include a certain number of neurons [13]. The neurons of the input layer and the hidden layer are fully connected in the network. It can also be seen in between the hidden layer and output layer (Figure 1). However, the neurons in the same layer do not communicate each other.

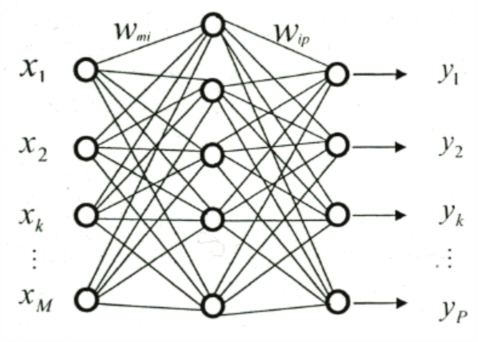

Figure 1. BP artificial neural network

A training sample is first input into the input layer in the neural network. Then the neurons of hide layers will calculate their values. Finally, the output layer will get the value (output) of this training sample. If the output has error by comparing desired result, the ANN will modify model parameters to minimize the error by using BP method [14]. The modified ANN will be trained by next training sample. The training process will stop when the model error reach a set value (such as 3%) [15].

B. ANN-CA for simulation of oil spill

The ANN-CA model can separate into two parts, training and simulation. However, these two parts use the same ANN. The detail process is listed as follow.

Each sample has *n* different properties, and a property is assigned to a neuron in the input layer. The sample can be expressed as follow:

$$X(k,t)=[x_1(k,t),x_2(k,t),...,x_i(k,t),...,x_n(k,t)]^{\mathrm{T}} \qquad (1)$$

where, $x_i(k,t)$ is the *i*th property at time *t* of cell (sample) *k*. T means transpose.

Before training, the training samples are usually needed to be standardized:

$$x_i^{'}(k,t)=(x_i\ (k,t)-\min)/(\max-\min) \qquad (2)$$

where, *max* and *min* is the max and min value in the *i*th property

After standardization, the training sample is input into the first layer – input layer. The neurons in the next layer (hide layer) will receive information (value) for the training sample, which can be expressed as follow:

$$net_j(k,t)=\sum_i w_{i,j}x_i^{'}(k,t) \qquad (3)$$

where, $net_j(k,t)$ is the received information of *j*th neuron in the hide layer. $w_{i,j}$ is the weight between input layer and hide layer.

The information will pass to the next layer – output layer. The neuron of output layer will get the information (probability, value) as follow:

$$P(k,t,l)=\sum_j w_{j,l}\frac{1}{1+e^{-net_j(k,t)}} \qquad (4)$$

where, $P(k,t,l)$ is the probability of become the type *l* at time *t*. $w_{j,l}$ is the weight between hide layer and output layer.

If the output has error by comparing desired result, the ANN will modify model parameters ($w_{i,j}$ and $w_{j,l}$) to minimize the error by using BP method. When the model error is controlled in a set value (such as 3%) , the training process will stop.

The trained ANN can be used to simulate the oil spill. However, there are many random factors in simulating real oil spill. It is believed stochastic disturbance term will help to obtain more plausible results [16]. The equation (5) can be change into following form.

$$P(k,t,l)=[1+(-\ln r)^{\alpha}]\times\sum_j w_{j,l}\frac{1}{1+e^{-net_j(k,t)}} \qquad (5)$$

This ANN will produce the change of probability of each cell. Then the model will decide the cell change into oil cell or not by comparing a predefined threshold. In order to control the iteration, a relatively high value is acceptable. And the change count is usually taken as the termination condition for simulation.

C. Study area and data processing

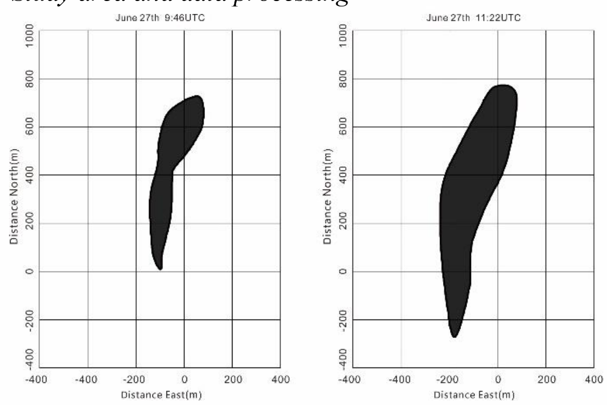

Figure 2. Slick contours obtained during the Marine Diesel experiment

The team members of "DeepSpill" project monitored the discharge of the mixture of diesel and natural gas at 27th June [9]. Two high-resolution images were collected in at 9:46 and 11:22. The spill area can be obtained using visual interpretation method (Figure 2). In our model, the image obtained at 9:46

will be input as the initial image, which is a key input value of the model. And the picture obtained at 11:22 will be used as the verification image of simulation results of CA model.

Five impact factors (Figure 3), including distance, currents, wind, salinity, and temperature, are chosen for simulating the surface behavior of oil spill. We obtain the center line of initial image (Figure 2a). The distance factor can be obtained by using Euclidean distance in the ArcGIS software (Figure 2b). The dispersion of the oil spill is considered as expanding outward from the center line of initial image (Figure 2a). The wind and current are also collected and projected at the dispersion direction (Figure 2c, d). Using the conductance temperature depth instrument (CTD), we also obtained the ocean temperature and salinity at the surface area (Figure 2e, f).

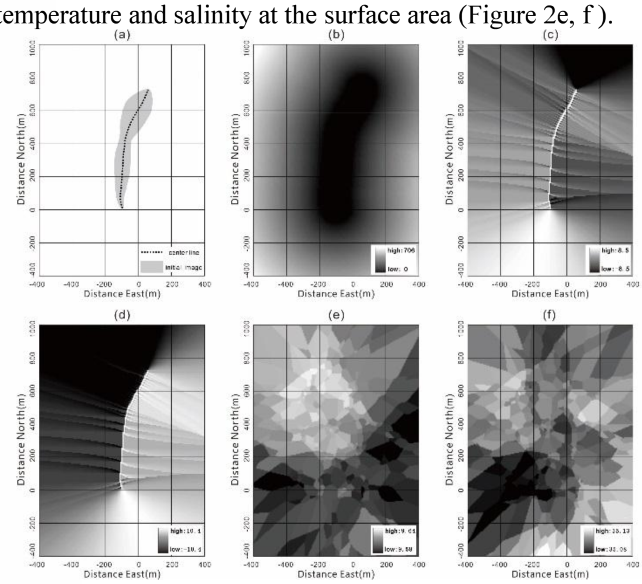

Figure 3. Oil Spill Impact Factor

(a) dispersion direction (b) proximity distance (c) currents
(d) wind (e) salinity (f) temperature

We select the samples for training and testing the parameters of CA model by using random sampling function in ArcGIS software. The samples were divided into two groups, training sample for mining rules and test sample for verifying the accuracy. The ratio of the samples are usually set to 20% of study area [17]. The parameters (weight) are trained by using training sample. We also verify the accuracy of the parameters by using the test samples. The accuracy of test samples is 97.6%, which can meet the requirements of geographical simulation by using CA models [17].

III. RESULT AND ANALYSIS

A. Simulation results

In this paper, an ANN-CA model for oil spills was constructed using C# programming. The simulation results (Figure 4) can be obtained by setting initial image (Figure 2), impact factors (Figure 3) and end condition.

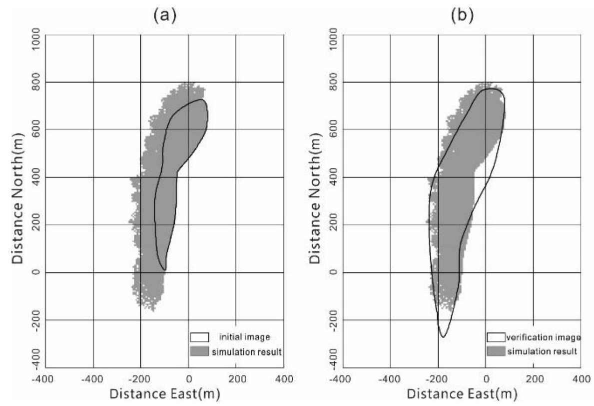

Figure 4. Simulation results

As can be seen from Figure 4a, the simulation results expand to the south and southwest, which is consistent with the verification image. The results hardly expand to the east because the eastern region is countercurrent area. The western area of simulation results fit better with verification image (Figure 4b). Better results also can be found in the northern and southern areas. However, model errors can be found in the eastern area because the neural network obtains less experience of eastward expansion in training. In the southern region, the oil areas expand about 200m to the south because of the effect of wind and current. However, the distance of moving south in simulation results is not more than 150m, which is shorter than the verification image. The model needs more iteration times to expand to a far distance. It is obvious that the more the iteration times, the more oil area will be obtained in the simulation results. However, the model will stop running when the oil areas are equal to the verification image.

TABLE I. ACCURACY AND KAPPA COEFFICIENT

		Simulation		
		oil	***non-oil***	***accuracy(%)***
Real	oil	3644	631	85.2
	non-oil	631	34005	98.2
	Total accuracy			96.8
	Kappa			0.834

The overall accuracy of simulation results can be obtained by using the confusion matrix (Table 1). It can be seen from the table that the accuracy is up to 98.1% in the non-oil areas. This value is associated with the ratio of oil area and non-oil area. If the study areas are large and the oil areas are only a small part of whole area, the value will be high. The accuracy of real and simulation results are in the oil area are 84.4% and 84.7% respectively. It is a high value for simulation results, which reflects precision in the non-oil area to oil area. The overall accuracy is 96.6%, the simulation results are reasonable for satisfied accuracy. Kappa coefficient is a commonly used indicator to evaluate the spatial pattern. It is found that the Kappa coefficient is as high as 0.826, that is, the simulation results are very similar to the verification image.

B. Analysis and discuss

1) The effect of sample ratio

This paper analyzes the effects of sampling ratio to the simulation results. The parameters in neural network will be different by using different training samples. Therefore, the simulation results will be different by using different parameters of neural network. In order to analyze the effect of sample ratio, we select the training sample using different ratio, including 1%, 5%, 10%, 15%, 20% and 30%. The training sample will be randomly selected 10 times using the same sampling ratio (eg.5%). The proposed CA model will repeat 10 times using the different selected training sample to verify the overlapping ratio (Figure 5). It can be found that the consistency of simulation results become well with an increasing proportion of the sample. It also means that the parameters in the neural network becomes more stable using more training samples. The simulation results of using sampling ratio of 1%, 5%, 10% and 15% have some drawback in consistency (Figure 5a, 5b, 5c. 5d). This showed that the parameters are not stable enough to express the conversion rule of oil spill using less training sample. However, better simulation results with high consistency are obtained using sampling ratio of 20% and 30%. It is recommended that the sampling ratio for training neural network is not less than 20%. However, high sampling ratio (eg. high than 50%) is not suggested for consuming more computation resources in training[15].

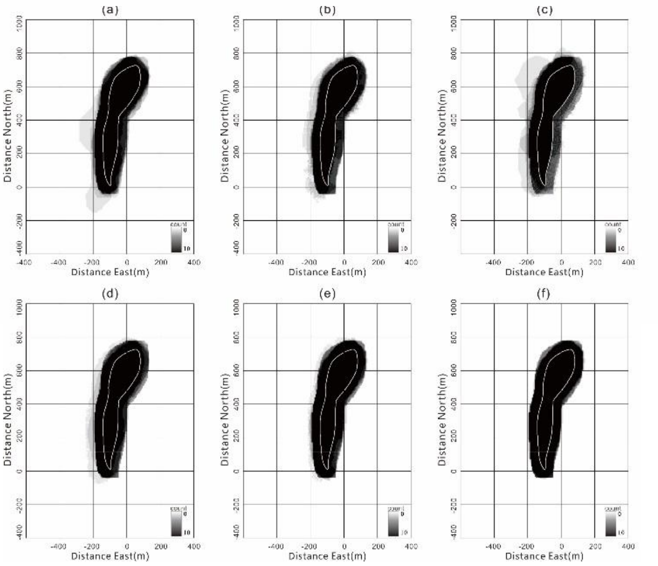

Figure 5. The ratio of training sample and simulation results

(a) 1% (b) 5% (c) 10% (d) 15% (e) 20% (f) 30%

2) *The effect of neuron count in hide layer*

The neuron count in hide layer may have some impact on the simulation results. The degree of fitting will be increase with the increasing neuron count in hide layer. However, it may get worse generalization ability because of over-fitting problem using too many neurons. In order to obtain better simulation results, it is necessary to set up an appropriate neuron count in hide layer. Experiments are carried out to study the impact on the simulation results using different neuron count in hide layer (Figure 6). It can be found that the consistency of simulation results increase then decrease with increasing neuron count in hide layer. The more neuron count in hide layer, the better the neural network can fit the training sample. Therefore, the simulation results become better with the increasing neuron count. However, the simulation results of consistency become worse as the neuron count higher than 10 in hide layer. This mainly because the neural network is so precisely represent the training samples that generalization ability of neural network decline. We also calculate the training time for using different neuron count. The training time for using 3, 5, 7, 10, 15 and 20 neurons in hide layer is 6.1s, 7.7s, 9.3s, 12.8s, 18.5s, and 24.0s, respectively. It can be found that the training time becoming longer for determine more parameters because of using more neuron count. Therefore, it is recommended that a suitable neuron count in hide layer should be selected to save training time and to obtain a neural network with better generalization.

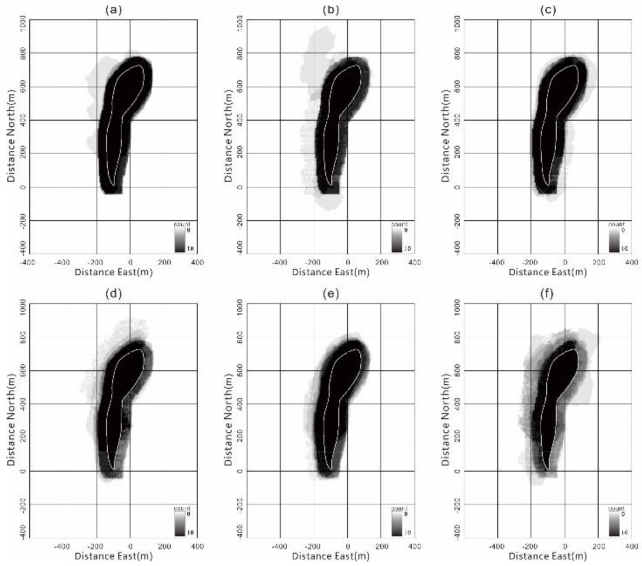

Figure 6. The count of neuron in hide layers and simulation results

(a) 3 (b) 5 (c) 7 (d) 10 (e) 15 (f) 20

IV. CONCLUSIONS

Model parameters are difficult to obtain in many traditional oil spill models. They cannot meet the requirements of rapid response for oil spill. Therefore, a new oil spill model--ANN oil spill CA model was established in this paper. This model can simulate the change process of oil spills by setting initial image, verification image, and impact factors. Experimental results show that the simulation results have a good performance with overall accuracy of 96.6% and Kappa coefficient of 0.826. It was also found that the consistency of simulation results is proportional to the ratio of training samples. However, the higher the ratio of training samples, the more computation is need in the ANN training. We have also studied the effect of neurons count in the hidden layer. Studies show that the consistency of simulation results becomes better with the increase of neurons number in the initial stage for good fitting rate of training sample. However, the consistency of simulation results gets worse for over-fitting of training sample in following stage.

The paper only simulates expansion behavior of oil spill, and further study will focus on how to simulate the disappearance of oil spill. More impact factors (such as temperature, salinity, light) will be involved for simulating the complex situation of oil spill.

ACKNOWLEDGMENT

The authors would like to express appreciations to Johansen for his valuable data, comments, and help.

REFERENCES

[1] W. Shi and M. Y. C. Pang, "Development of Voronoi-based cellular automata -an integrated dynamic model for Geographical Information Systems," International Journal of Geographical Information Science, vol. 14, pp. 455-474, 2000.

[2] C. He, N. Okada, Q. Zhang, P. Shi, and J. Zhang, "Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China," Applied Geography, vol. 26, pp. 323–345, 2006.

[3] X. LI and X. Liu, "An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region," International Journal of Geographical Information Science, vol. 20, pp. 1109-1136, 2006.

[4] L. Mateus, Rocha, W. Hordijk, and I. Group, "Material Representations: From the Genetic Code to the Evolution of Cellular Automata," Artificial Life, vol. 11, pp. 189 - 214, 2005.

[5] M. Aljoufie, M. Zuidgeest, M. Brussel, J. v. Vliet, and M. v. Maarseveen, "A cellular automata-based land use and transport interaction model applied to Jeddah, Saudi Arabia," Landscape and Urban Planning, vol. 112, pp. 89-99, 2013.

[6] Y. H. Zhang, X. Li, X. P. Liu, and J. G. Qiao, "The CA model based on data assimilation," Journal of Remote Sensing, vol. 15, pp. 475–491, 2011.

[7] X. Chao, N. J. Shankar, M. Asce, S. S. Y. Wang, and F. Asce, "Development and application of oil spill model for Singapore coastal waters," Journal of Hydraulic Engineering, vol. 129, pp. 495-503, 2003.

[8] K. M. Gamzaev, "Modeling the spread of an oil slick on the sea surface," Journal of applied mechanics and technical physics, vol. 50, pp. 466-469, 2009.

[9] Ø. Johansen, H. Rye, and C. Cooper, "DeepSpill—field study of a simulated oil and gas blowout in deep water," Spill Science & Technology Bulletin, vol. 8, pp. 433-443, 2003.

[10] I. Karafyllidis, "A model for the prediction of oil slick movement and spreading using cellular automata," Environment international, vol. 23, pp. 839-850, 1997.

[11] Z. Rušinović and N. Bogunović, "Cellular automata based model for the prediction of oil slicks behavior," in 26th International Conference on Information Technology Interfaces, 2006.

[12] J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, and P. Freedson, "An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer," Journal of Applied Physiology, vol. 107, pp. 1300-1307, 2009.

[13] J. D. Wu, N. Li, H. J. Yang, and C. H. Li, "Risk evaluation ofheavy snow disasters using BP artificial neural network: the case of Xilingol in Inner Mongolia," Stochastic Environmental Research and Risk Assessment, vol. 22, pp. 719-725, 2008.

[14] K. Huang, L. Dai, and S. Huang, "Wind Prediction Based on Improved BP Artificial Neural Network in Wind Farm," in Electrical and Control Engineering (ICECE), 2010 International Conference on, 2010, pp. 2548 - 2551.

[15] X. Li and A. G.-O. Yeh, "Neural-network-based cellular automata for simulating multiple landuse changes using GIS," International Journal of Geographical Information Science, vol. 16, pp. 323-343, 2002.

[16] R. White and G. Engelen, "Cellular Automata and Fractal Urban Form: A Cellular Modelling Approach to the Evolution of Urban Land Use Patterns," Environment and Planning A vol. 25, pp. 1175–1199, 1993.

[17] X. Li and A. G.-O. Yeh, "Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta," International Journal of Remote Sensing, vol. 19, pp. 1501-1518, 1998.