Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 14	No. 3	pp. 727-732	2015

Original Research Paper

Water Quality Prediction Based on BP Neural Network at Dahuofang Reservoir, China

Lingling Ma, Linfei Zhou and Tieliang Wang†
College of Water Resource, Shenyang Agricultural University, Shenyang, Liaoning, 110866, China
†Corresponding author: Tieliang Wang

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 8-4-2015
Accepted: 5-5-2015

Key Words:
BP neural network
Water quality prediction
Simulation
Dahuofang reservoir

ABSTRACT

To ensure the safety of drinking water, understanding the trends of water quality in water resource and to provide a scientific basis for water quality management, a three-layer BP neural network is selected to simulate and predict six water quality indicators of the outbound of Dahuofang Reservoir. The six water quality indicators are dissolved oxygen, five days' biochemical oxygen demand, permanganate index, ammonia nitrogen, total nitrogen and total phosphorus. Training the model with water quality data from 2005 to 2011, Levenberg-Marguardt optimization algorithm is adopted to train samples. After reaching the error requirement, simulate the model with the water quality monitoring data in 2012 and test the model accuracy. Simulation results show that the accuracy of the model prediction is higher in 2012. It is proved that this model can be used to predict water quality of the outbound mouth in Fushun section, and the model provides a theoretical basis for improving the water quality of the reservoir area and can be used to guide the actual water quality management.

INTRODUCTION

Water quality tends to deteriorate gradually with human interventions, such as land use change (Wagner et al. 2013), hydrological alterations (Arias et al. 2014), application of chemical fertilizers and pesticides (Matthews et al. 2012, Álvarez-Romero et al. 2014), which cause series of environmental problems, such as soil erosion (Keesstra et al. 2013), shortage and deterioration of drinking water (Khan et al. 2013), and degeneration of aquatic ecosystems (Ryan et al. 2013). With water pollution becoming an increasingly serious problem, it has attracted much attention from the public and the government. Therefore, integrated research on water quality prediction is one of the most important research issues.

Dahuofang reservoir is located 18km east of Fushun city, Liaoning Province, and is a valley reservoir, whose main role is controlling flood, irrigation, industrial water supply and drinking water. With the accomplishment of water conveyance project of Dahuofang reservoir, it has become a water supply source for 23 million population in Liaoning Province, and is one of the nine important water source areas in China. In recent years, with the rapid development of the local economy, pollution load in the reservoir increases year by year. Higher use of chemical fertilizers and pesticides pollutes ground and surface water, along with surface runoff and soil erosion. Livestock manure and garbage without effective treatment timely enter into the reservoir or river with the rain and pollute the water source seriously. Dahuofang reservoir plays a key role in the development of national economy and social stability in Liaoning Province. Systematic researching and grasping the changes of water quality of Dahuofang reservoir will play an important part in controlling and improving the water quality and ecological environment, and will make the indicators of water quality and ecology meet the national type II standard of surface water in China.

The water quality prediction model can be divided into mechanistic prediction model and non-mechanistic prediction model. Mechanistic prediction model is usually very complex, the calculation process requires more basic information and parameters, while our nation relatively lacks the basic information and data of the majority of our rivers, lakes and reservoirs, which limits the practical application of mechanistic prediction model for water quality forecasting (Fan et al. 2010). While non-mechanistic prediction model is a "black box" approach, establishing a water quality model through mathematics, has an ideal forecasting effect, and is widely used in simulation and prediction of water quality. The common ones of non-mechanistic prediction model are artificial neural networks (Perera et al. 2014, Sarkar et al. 2015), gray theory prediction method (Wei et al. 2013), the time series method (Liang 2014), Markov method (Wöhling et al. 2011, Tong et al. 2012).

All things considered, an artificial neural network method for the prediction of water quality will be selected in this study.

MATERIALS AND METHODS

Study area: Dahuofang reservoir is situated at east longitude 123°04'28"to 124°27'46"and north latitude 42°21'10" to 42°04'01", and is located in the eastern part of Fushun City, Liaoning Province, middle and upper area of Hun River, 18km away from Fushun city. The reservoir was completed and put to use in 1958. All basins are distributed in three counties, which are Qingyuan, Xinbin and Fushun in Fushun city. The reservoir controls total area of 5437km^2 of all basins. The maximum capacity is 2.268 billion m^3. Dahuofang reservoir accepts the surface water from Hun River, She River and Suzi River. Hun river is originated in Qingyuan county, from east to west throughout the whole Fushun city. Hun river is 169km long above the dam site, the river slope is 2.8 ‰, and flowing into Dahuofang reservoir in Beizamu. Suzi river originates in Xinbin county, river length is 132km, the river slope is 2.3 ‰, and flowing into Dahuofang reservoir in Gulou. She river originates in Fushun county and is 46km long, flowing into the reservoir in Taigou, and the river slope is 7 ‰.

There are 7 monitoring sections in Dahuofang reservoir, respectively, warehousing mouth: Taigou (She River), Gulou (Suzi River), Beizamu (Hun River); the reservoir area: Hun37 (middle of the reservoir), Hun73 (tail of the reservoir), Hun7 (front of the reservoir); outbound mouth: outbound mouth in Fushun (front of the dam). The sectional distribution is shown in Fig. 1. Sampling points of each section of the reservoir are given in Table 1.

BP neural network model: Artificial neural network belongs to the "black box" model, one of the most widely used is BP (Back Propagation) neural network. The team of scientists led by Rumelhart and McCelland raised BP (Back-Propagation) neural network in 1986, which is a multilayer forward neural network that is trained by the error back-propagation algorithm. Currently, in the practical application of artificial neural networks, the vast majority use of the neural network models is BP network and its variations. It is the core part of forward networks, reflecting the essence of the artificial neural networks.

BP network is a kind of three-layer or more neural network, composed of an input layer, hidden layer and an output layer. A full connection between layers and there is no connection between neurons of each layer, as shown in Fig. 2. Its main characteristic is signal forward propagation and error back propagation. In the forward propagation, information inputs from the input layer, calculates layer by layer in the hidden layer, and then transports to the output layer. State of neurons in each layer only affects the status of next-layer neurons. If the output layer does not get the desired output, calculate the error variance of the output layer, and then turn to back propagation. The error signal propagates through back-propagation along the original

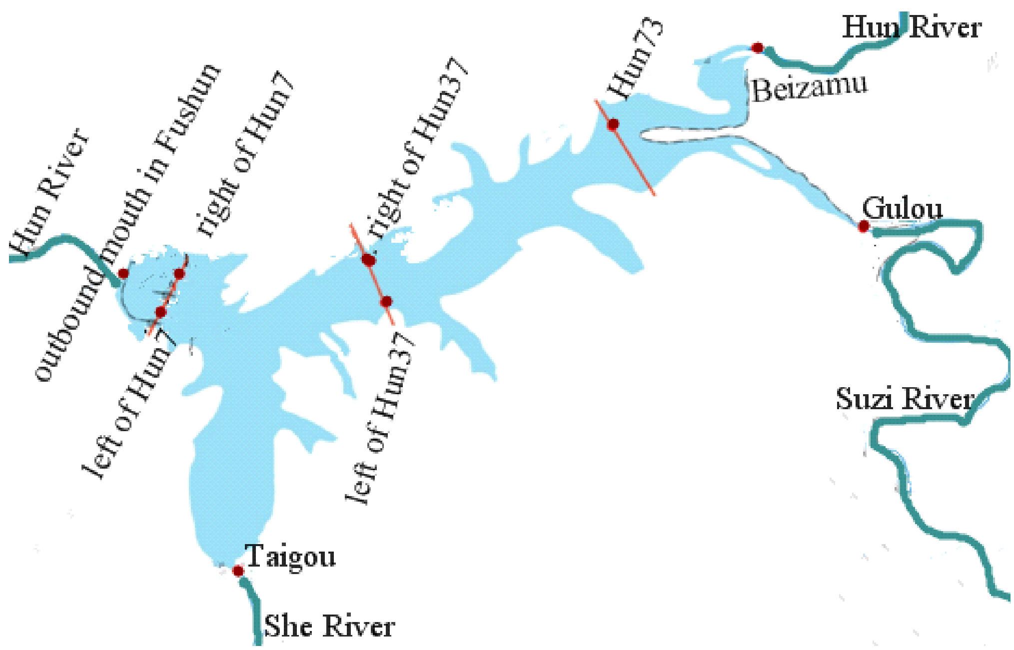

Fig. 1: Dahuofang reservoir monitoring section location map.

Table 1: Sampling section layout of Dahuofang resevoir.

	Warehousing mouth			Reservoir area			Outbound
	Hun River	SuziRiver	She River	Tail of the reservoir	Middle of the reservoir	Front of the reservoir	
Section	Beizamu	Gulou	Taigou	Hun73	Hun37	Hun7	Outbound mouth in Fushun
Vertical line	1	1	1	1	2	2	1
Sampling point	1	1	1	1	6	6	1

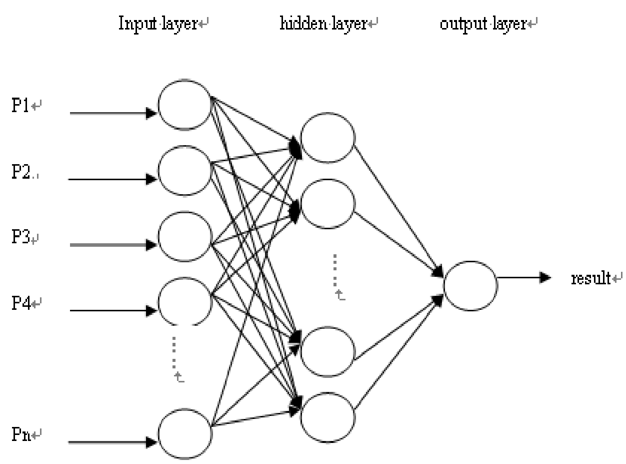

Fig. 2: BP neural network topology structure.

connection pathway. The network modifies the neurons' weight of each layer according to the error change value, repeated iterative operation, until the error is reduced to the desired target. The typical BP network has three layers, which contains one hidden layer. This structure not only meets the accuracy requirements, but also can accelerate the convergence of the network and reduce training time. The BP neural network is applied to predict the water quality by many scholars and better results are obtained (Yan et al. 2014, Tong et al. 2014, Mallesh 2011, Xue et al. 2013).

Steps to Establish BP Neural Network

1. Select the training samples and pretreat. Determine input layer and output layer neurons, and normalize the samples, which can avoid the differences of magnitude to affect network precision.
2. Establish BP network and train the network. Select the number of network layers and the number of hidden layer neurons. The network format is: net = newff (PR, [S1, S2…SN],{TF1, TF2…TFN}, BTF, BLF, PF). PR is the matrix of ranges of network input vector, [S1, S2…SN] is the neuron number of hidden layer and output layer, {TF1, TF2…TFN} is the transfer function of hidden layer and output layer, BTF is the network training function, BLF is weight learning function, PF is performance function.
3. Simulate the model and determine whether can be applied to the water quality prediction actually.

Data source: This research adopts the samples for water quality data of the period 2005-2012. After comparing the evaluation criteria and actual indicators, six representative indices are selected: dissolved oxygen (DO), five days biochemical oxygen demand (BOD_5), permanganate index (COD_{Mn}), ammonia nitrogen (NH_3-N), total nitrogen (TN), and total phosphorus (TP). Data are the annual average of monitoring section.

Selection of training samples: Select the water pollution index data from 2005 to 2011 of Taigou (She River), Gulou (Suzi River), Beizamu (Hun River); the reservoir area: Hun37 (middle of the reservoir), Hun73 (tail of the reservoir), Hun7 (front of the reservoir) as input samples to establish a network, which are training samples. Select the index data of outbound mouth in Fushun (front of the dam) as output layer neurons. Simulate and predict DO, BOD_5, COD_{Mn}, NH_3-N, TN and TP respectively. Select the monitoring data of 2012 as test samples to verify the accuracy of the network.

Pretreatment of training samples: Different pollution index data have different magnitudes. Carry out normalization processing of training samples using prestd function in MATLAB.

Establish network structure: Typically, neural network does not need to be too complicated. Complex networks often require more parameters. If too many parameters are estimated, the accuracy of network will decline naturally, network forecasting accuracy will decline too. According to the universal approximation theorem "a three-layer BP network with a hidden layer, as long as there are sufficient number of nodes, it can approximate any continuous function on a bounded region with arbitrary precision", so a three-layer BP network with a hidden layer is selected.

Table 2: The results of water simulation and prediction.

Year		DO	COD_{Mn}	BOD_5	NH_3-N	TN	TP
2005	Actual value ($mg.L^{-1}$)	8.1000	2.7000	0.8000	0.1000	1.4200	0.1000
	Predicted value ($mg.L^{-1}$)	8.1000	2.6962	0.8006	0.0997	1.4198	0.1000
	Absolute error	0.0000	0.0038	0.0006	0.0003	0.0002	0.0000
2006	Actual value ($mg.L^{-1}$)	8.5000	3.1000	1.0000	0.1800	2.1100	0.0150
	Predicted value ($mg.L^{-1}$)	8.5000	3.1002	1.0033	0.1796	2.1100	0.0150
	Absolute error	0.0000	0.0002	0.0033	0.0004	0.0000	0.0000
2007	Actual value ($mg.L^{-1}$)	6.7000	2.8000	1.2500	0.2300	1.6100	0.0100
	Predicted value ($mg.L^{-1}$)	6.7000	2.7939	1.2500	0.2299	1.6099	0.0100
	Absolute error	0.0000	0.0061	0.0000	0.0001	0.0001	0.0000
2008	Actual value ($mg.L^{-1}$)	8.0500	3.3000	1.6000	0.1600	1.5800	0.0150
	Predicted value ($mg.L^{-1}$)	8.0500	3.2999	1.6000	0.1599	1.5800	0.0150
	Absolute error	0.0000	0.0001	0.0000	0.0001	0.0000	0.0000
2009	Actual value ($mg.L^{-1}$)	8.6500	2.6000	1.4000	0.1000	1.8700	0.0150
	Predicted value ($mg.L^{-1}$)	8.6501	2.6012	1.4000	0.0997	1.8700	0.0150
	Absolute error	0.0001	0.0012	0.0000	0.0003	0.0000	0.0000
2010	Actual value ($mg.L^{-1}$)	8.2500	2.5000	1.1000	0.2100	1.1500	0.0100
	Predicted value ($mg.L^{-1}$)	8.2501	2.4991	1.1000	0.2100	1.1500	0.0100
	Absolute error	0.0001	0.0009	0.0000	0.0000	0.0000	0.0000
2011	Actual value ($mg.L^{-1}$)	9.0000	2.1000	1.2000	0.1200	1.4500	0.0100
	Predicted value ($mg.L^{-1}$)	9.0006	2.1007	1.2000	0.1202	1.4500	0.0100
	Absolute error	0.0006	0.0007	0.0000	-0.0002	0.0000	0.0000

Table 3: The simulation results of water simulation and predietion.

	DO	COD_{Mn}	BOD_5	NH_3-N	TN	TP
Actual value ($mg.L^{-1}$)	10.0000	2.0000	1.3000	0.1350	0.9800	0.0100
Predicted value ($mg.L^{-1}$)	9.9225	2.1814	1.4010	0.1374	1.0620	0.0103
Absolute error	0.0775	-0.1814	-0.1010	-0.0024	-0.0820	-0.0003
Relative error (%)	0.7750	-9.0700	-7.7690	-1.7778	-8.3670	-3.0000

Determining the number of nodes in the hidden layer is quite a complex issue. If the number of nodes are too less, network training will be insufficient and the fault-tolerance performance is poor. If the number of nodes are excessive, learning time of the network will be extended and it can not guarantee an optimal error. At present, there are no mature and complete theoretical methods to determine the hidden layer nodes of the network. The trial and error approach is adopted to determine the number of hidden layer nodes. Firstly, determine the approximate range based on empirical formula $l<\sqrt{(m+n)}+a$, and then determine the optimal number of nodes by trial and error approach. Where l is the number of hidden layer nodes, m is the number of output layer nodes, n is the number of input layer nodes, a is the constant between 0-10. When the number of hidden layer nodes is 3, training steps are less and overall error is small. Therefore, determine the number of nodes of the hidden layer as 3.

Select the network parameters: Select the tansig function as the hidden layer transfer function, which is a hyperbolic tangent S-type transfer function, and will achieve arbitrary nonlinear mapping between inputs and outputs. Select purelin as the output layer transfer function, which is a linear function. Through the comparison of the performance of several learning algorithms, it is found that Levenberg-Marguardt algorithm convergence is satisfactory and has a small error. So trainlm is selected as the network training function, which is Levenberg-Marquardt algorithm based on numerical optimization (Referred LM algorithm). This algorithm is an improved method of the original BP algorithm. It has the fastest convergence rate for the medium-sized BP neural network, which not only increases the reliability of the algorithm but also improves the learning rate of the network. LM algorithm is a fast algorithm with standard numerical optimization techniques. It is a combination of gradient descent algorithm and Gauss-Newton method, which can also be said to be the modified form of Gauss-Newton method. It has both local convergence performance of Gauss-Newton method and global characteristics of the gradient descent method. With the aid of the approximate second derivative, the LM algorithm is more efficient than the gradient method. Select learngdm function as the weight learning function of the network. It is a decreasing function using the gradient descent method with batch process and adding a momentum item and which is a feedforward neural network training method. It not only has faster convergence but also introduces a momentum term, which effectively avoids local minima problems in network training. Select mean square error performance analysis function mse as network performance function. Set the display frequency of training process at 10 times, the initial learning rate at 0.01, the maximum training cycles at 100 times and anticipation error of network training at 1e-4.

Train the network: Train the network based on MATLAB 7.0.1. Its implementation code is as follows:

```
net = newff(minmax(p),[3,1],{'tansig', 'purelin'},'
trainlm','learngdm','mse');
net.trainParam.show=10;
net.trainParam.lr=0.01;
net.trainParam.epochs=100;
net.trainParam.goal=1e-4;
net=train(net,p,tt);
```

RESULTS AND DISCUSSION

Predicting the main pollutants of 2005-2011 with water quality simulation model, which has been built, Table 2 shows the predictive value of DO, COD_{Mn}, BOD_{50}, NH_3-N, TN, TP at the outbound mouth in Fushun section and the absolute error of predicted value and the actual value.

Table 2 shows that the effect of simulation and prediction of water quality monitoring indicators is satisfactory, absolute errors are all less than 0.01. Therefore, simulation training of the network is successful.

Verify the water quality simulation and prediction model: Weights and the threshold value of network have been determined after being trained. Now, only need to validate the trained model to determine whether the model can be used to practice and achieve the intended purpose. We need to take the test samples into the network to validate the model and calculate the network output. Analyse error of the output value and the actual monitoring value, if the network accuracy meets the requirement, the model can be used to predict water quality. The statement based on the MATLAB is:

```
[y,meanp,stdp]=prestd(y)
y1=y'
y2=sim(net,y1)
y3=poststd(y2,meanp,stdp);
```

Where, sim is the simulation function, prestd is normalization function, and poststd is anti-normalization function. The test results are given in Table 3.

Simulation results show that the simulation effect is better and the relative error is between 0.7750% and 9.0700%. The network has been relatively stable after training, model prediction accuracy meets the requirements and it can be used to forecast the water quality of outbound mouth in Fushun of Dahuofang reservoir.

CONCLUSIONS

1. In this study, the BP neural network method was adopted to predict the outbound water quality of Dahuofang reservoir in 2012. The results show that the predictive accuracy of the model is high, relative error is small, and is between 0.775% and 9.07%, which is less than 10%. The predicted results are objective and reasonable, and it has a strong nonlinear predictive ability. Thus, the model can be applied to forecast these six indicators of any section. Only need to know the actual value of certain indicator of other six sections, we can then calculate the predicted values of the section to be predicted.
2. Predict the water quality with application of neural network, through a simple multiple complex of nonlinear function to achieve nonlinear accurate mapping, which has strong adaptive ability and generalization ability and provides a new research method on water quality prediction. With the help of MATLAB, it is easy to operate the prediction process without manual calculation. It is convenient for application, only need to enter the actual monitoring data, and can obtain predictive values.
3. BP neural network method is adopted to forecast water quality of outbound and which will be targeted to improve the water environment in the reservoir area and ensure the safety of drinking water. It can also provide a realistic approach to environmental control of large-scale projects.

ACKNOWLEDGEMENTS

This work was financially supported by the National Science Foundation of China (50879046) and the Agriculture Research Projects of Science and Technology Office of Liaoning, China (2012212001).

REFERENCES

Álvarez-Romero, J. G., Wilkinson, S. N., Pressey, R. L., Ban, N. C., Kool, J. and Brodie, J. 2014. Modeling catchment nutrients and sediment loads to inform regional management of water quality in coastal-marine ecosystems: A comparison of two approaches. Journal of Environmental Management, 146: 164-178.

Arias, M. E., Piman, T., Lauri, H., Cochrane, T.A. and Kummu, M. 2014. Dams on Mekong tributaries as significant contributors of hydrological alterations to the Tonle Sap floodplain in Cambodia. Hydrology and Earth System Sciences, 18(12): 5303-5315.

Fan, M. and Gu, Z. L. 2010. Water quality model progress and development trend. Shanghai Environmental Sciences, 29(6): 266-274 (in Chinese).

Keesstra, S. D., Arricibita, A. M., Lassu, T. and Seeger, M. 2013. Rainfall intensity, aggregate stability, shear stress resistance as parameters to evaluate soil erosion by water: An experimental study. In: Proceedings of the 8th IAG International Conference on Geomorphology, Paris, France, 8, pp. 27-31

Khan, S., Shahnaz, M., Jehan, N., Rehman, S., Shah, M. T. and Din, I. 2013. Drinking water quality and human health risk in Charsadda district, Pakistan. Journal of Cleaner Production, 60: 93-101.

Liang, Y. Z., Liu, Y. and Sheng, H. 2014. Identification of water quality trends in Lake Dianchi using an integrated STL and 2, SI approach, Acta Scientiae Circumstantiae, 34(3): 754-762 (in Chinese).

Mallesh, T. V., Prakash, S. M., Kumar, L. P. and Jayaramappa, N. 2011. Evaluation of bacteriological parameters in water using artificial neural network. Nature, Environment and Pollution Technology, 10(1): 159-166.

Matthews, R. D., Anderson, D., Burden, D., Johns, G., Janicki, A., Zarbock, H. and Pribble, R. 2012. Linking receiving water impacts to sources and to water quality decisions: Using nutrients as an initial case study. Proceedings of the Water Environment Federation, 11: 5025-5039.

Perera, K. C., Western, A. W., Nawarathna, B. and George, B. 2014. Forecasting daily reference evapotranspiration for Australia using numerical weather prediction outputs. Agricultural and Forest Meteorology, 194: 50-63.

Ryan, M. E., Johnson, J. R., Fitzpatrick, B. M., Lowenstine, L. J., Picco, A. M. and Shaffer, H. B. 2013. Lethal effects of water quality on threatened California salamanders but not on co-occurring hybrid salamanders. Conservation Biology, 27(1): 95-102.

Sarkar, A. and Pandey, P. 2015. River water quality modelling using artificial neural network technique. Aquatic Procedia, 4: 1070-1077.

Tong, L., Wan, H. and Guan, G. 2012. Prediction model for moisture content of purified terephthalic acid recovery system based on BP neural network. Journal of Nanjing University of Technology (Natural Science Edition), 34(3): 70-73.

Tong, S. T., Sun, Y., Ranatunga, T., He, J. and Yang, Y. J. 2012. Predicting plausible impacts of sets of climate and land use change scenarios on water resources. Applied Geography, 32(2): 477-489.

Wagner, P.D., Kumar, S. and Schneider, K. 2013. An assessment of land use change impacts on the water resources of the Mula and Mutha rivers catchment upstream of Pune, India. Hydrology and Earth System Sciences, 17(6): 2233-2246.

Wei, Z. K., Jiang, S. Y. and Li, S. D. 2013. Application of grey theory in water quality prediction of sudden cadmium pollution in Longjiang River. Journal of Water Resources, Water Engineering, 24(3): 135-141 (in Chinese).

Wöhling, T. and Vrugt, J. A. 2011. Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. Water Resources Research, 47(4), DOI: 10.1029/2010WR009265.

Xue, M. 2013. A novel water quality assessment method based on combination of BP neural network model and fuzzy system. Journal of Computers, 8(6): 1587-1593.

Yan, J. X., Yu, L. J., Mao, W. W. and Cao, S. Q. 2014. Study on prediction model of dissolved oxygen about water quality monitoring system based on BP neural network. Advanced Materials Research, 912: 1407-1411.