2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Applying Computational Intelligence Methods to Modeling and Predicting Common Bean Germination Rates

A. Bianconi, M. J. Watts, Y. Huang, A. B. S. Serapião, J. S. Govone, X. Mi, G. Habermann and A. Ferrarini

Abstract— **The relationship between seed germination rate and environmental temperature is complex. This study assessed the effectiveness of multi-layer perceptron (MLP) and Particle Swarm Optimization (PSO) techniques in modeling and predicting the germination rate of two common bean cultivars as a function of distinct temperatures. MLP was utilized to model the germination rate of the cultivars and PSO was employed to determine the optimum temperatures at which the beans germinate most rapidly. The outcomes derived from implementing the MLP were compared with those obtained by means of a traditional statistical method. The MLP provided more accurate results than the conventional statistical regression in predicting germination rate values regarding the two common bean cultivars. The optimum germination rate values derived from implementing the PSO model were more accurate than those obtained by using the conventional quadratic regression.**

I. INTRODUCTION

The common bean (*Phaseolus vulgaris* L.) is an important cash and subsistence crop in several parts of the world [11, 27, 8,]. Therefore, the understanding of every developmental stage of this species is essential for implementing feasible production techniques [2] and studies of germination underpin studies concerning more advanced developmental phases [2, 20].

The germination of seeds has been described as both a critical period and a complex phenomenon [15, 2, 25, 31, 3]. This is because life-history traits would not be expressed if plants were not able to survive in the germination phase [6]. Of the factors that affect the germination process, temperature has a primary influence on seed germination, affecting the rate of germination [12, 1, 30, 19], where higher temperatures might well cause germination rate values to rise [30, 29]. Therefore, the intrinsic characteristics of seeds that govern their germination in various temperatures should be taken into consideration when analyzing the biology of seeds [12, 26].

The relationship between germination rate and temperature could be regarded as a simple relationship, because most papers on seed germination have not utilized complex models in predicting germination rate as a function of different temperatures. That is, most papers on seed biology have used classical uni- or multivariate statistical procedures for analyzing seed germination (e.g., [12, 1, 15, 25, 2]). However, the relationship between germination rate and temperature can be complex [2, 10], and simple statistical methods might not be able to reveal the complexity of this phenomenon.

If more complex analytical tools had been widely used in analyzing germination rate as a function of temperature, and if the outcomes of these tools were the same as those derived from conventional statistical techniques, then it would have been fair to conclude that the relationship between germination rate and temperature might well be a simple phenomenon. However, even if a relationship between germination rate and temperature were simple, obtaining precise and accurate models that could describe a large proportion of the total variance could never be regarded as a simple modeling procedure. Moreover, the common bean is a profitable agricultural product [27, 8, 2], and modeling the germination rate of this crop species requires precision and accuracy because of its commercial value. Standard statistical methods could model this relationship (e.g., [12,1]), but they may be limited by both the assumption of linearity and other statistical assumptions. Soft computing techniques such as artificial neural networks (ANNs) provide an effective approach for modeling nonlinear processes. ANNs have previously been applied to several problems in plant biology [7, 18, 21, 9, 32, 14] and ecology [33]. For example, [13] used neural modeling to analyze the percentage germination of seeds during grain drying processes, and [5] implemented some ANN models to predict the accumulated emergence of a weed species. However, neural modeling techniques have not been widely applied to analyzing seed germination. Indeed, the current paper represents the first application of neural modeling to the modeling and prediction of the germination rate (germination speed) of a crop species.

The objectives of this research were: 1) to assess the effectiveness of a MLP neural network in modeling the germination rate of two common bean cultivars as a function of distinct temperatures under laboratory conditions; 2) to utilize the Particle Swarm Optimization (PSO) algorithm for

A. Bianconi is with the International Academy of Ecology and Environmental Sciences, Hong Kong. China. e-mail drebianconi@yahoo.com.br

M. J. Watts is with the Information Technology Programme, Auckland Institute of Studies, Auckland, New Zealand, email mjwatts@ieee.org (Corresponding author)

Y. Huang is with the United States Department of Agriculture (USDA), Agricultural Research Service, Crop Production Systems Research Unit, Mississippi, USA

A. B. S. Serapião and J. S. Govone are with IGCE-DEMAC-Unesp, Rio Claro-SP, Brazil

X. Mi is with the State Key Laboratory of Vegetation and Environmental Change, Institute of Botany Chinese Academy of Sciences, Beijing, China

G. Habermann is with Instituto de Biociências, Unesp, Rio Claro-SP, Brazil

A. Ferrarini is with The University of Parma, Parma, Italy.

The Fundação de Amparo à Pesquisa do Estado de São Paulo provided A. Bianconi with financial support during the collection of the data sets used in this paper.

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

determining the optimum temperatures at which the two bean cultivars germinate most rapidly; and 3) to compare the results from neural network models and PSO with those derived from a standard polynomial regression method.

II. MATERIALS AND METHODS

A. Laboratory experiments

'Carioca' and 'Iapar 81' were the two Brazilian bean cultivars used in this work. The laboratory experiments used a temperature gradient block [17] providing 25 distinct temperatures. The sample size was equal to 125 for both cultivars (five replicates per temperature). Each group of five replicates was exposed to one of the 25 constant temperatures, and the same temperatures were used for both cultivars.

B. Variables

Temperature was the explanatory variable (input) and germination rate (germination speed) was the dependent variable (output). The germination rate utilised in the present work represents the inverse of the mean time of germination (i.e. mean time^{-1}, measured in hours^{-1} in this study) and can be written as:

$$\text{mean time} = \left[\left(\sum_{i=1}^{k} n_i T_i \right) \Big/ \left(\sum_{i=1}^{k} n_i \right) \right] \qquad (1)$$

in which T_i is the elapsed time from the start of the experiment to the k_{th} observation (in hours); n_i is the number of seeds germinated at time i. Additional details concerning this germination index can be found in [28].

C. Statistical model

A traditional polynomial regression method was used as a benchmark against which the outcomes of the neural models were compared. Polynomial regression models are special cases of the linear regression model [16]. The following is the polynomial regression model utilized in the current investigation:

$$\hat{y}_i = b_0 + b_1 x_i + b_2 x_i^2 \qquad (2)$$

where $\hat{y}_i$ denotes the response (the predicted germination rate) in the i_{th} trial; x_i represents the value of the predictor variable (temperature) in the i_{th} trial; and b_0, b_1, and b_2 represent the parameters of the model.

The performance derived from the MLP network was compared with that derived from the polynomial regression by means of R^2 values (coefficient of determination). This coefficient constitutes a simple and effective means of comparing the actual values (observed values) with those derived from the predictive models (estimated values). In addition, root mean square errors (RMSE) were utilized in evaluating the performance of the models. Additional information on these well-known statistics can be found in [16].

D. Modeling procedures for analyzing germination rate

The MLP had five neurons in the hidden layer. The number of neurons in this layer was determined by assessing which configuration yielded the lowest global training error. The activation functions used at each level of the network were sigmoid for the hidden layer neurons and linear for the output neuron, and standard backpropagation was used for training. Each bean cultivar was modeled separately.

To train the MLP, the entire data set (i.e. 125 examples for each bean cultivar) was randomly separated into two subsets: 60% of the examples were used as the training subset (75 examples), and 40% were used as the test subset (50 examples). The training and test subsets for the statistical regression model were the same as those used for the MLP network.

The 'learned' neural function enabled us to achieve two aims: it was used as a predictive tool, estimating germination rate values at any temperature; and it functioned as an objective function (meta-heuristic) of an optimization algorithm that was capable of providing the neural function with near-optimum solutions, after which the Particle Swarm Optimization (PSO, see [36] and [23] for details about this neural model) was used to obtain the near-optimum germination rate value for each cultivar.

Once the PSO had determined the optimum germination rate, it also provided the related temperature. In addition, model parameters were estimated with the purpose of maximizing the output function that was derived from the model inputs. Thus, PSO obtained the temperature that produced the maximum germination rate value (i.e. the optimum germination rate) from the implemented neural network.

Distinct results were obtained by the optimization process using the parameters employed in the PSO algorithm. This optimization process was carried out several times, using the following parameters: swarm size 15; learning coefficients were equal to 2; inertial weight (w) fixed at 1; and the maximum number of epochs was restricted to 30.

In order to compare the optimum germination values derived from the PSO model with those obtained by using the traditional quadratic regression model, the following equation was used [35]:

$$x_0 = (-b_1)/(2b_2) \qquad (3)$$

in which x_0 represents the optimum temperature (independent variable); and b_1 and b_2 are the regression coefficients (see Equation 2). The optimum germination rate was obtained by substituting the estimated x_0 for the x_is in the estimated quadratic model.

E. Akaike's Information Criterion

The larger the number of parameters in a model, the better the fit of the model. However, the main drawback to an overparameterized model is its inability to generalize.

Therefore, we employed Akaike's Information Criterion (AIC), which compromises between goodness of fit and parsimony in the number of parameters [34]. In other words, AIC penalizes models containing large numbers of parameters [34, 16]. It can be written as:

$$AIC=n\ln R+2p \qquad (4)$$

where n represents the number of training or test subset values. Both the training and test data sets used in the statistical model were exactly the same as those used in implementing the MLP. R represents the residual sum of squares, and p is the number of parameters in the model. Three parameters were used for the quadratic regression, and 10 for the MLP. Normally, a lower AIC value indicates a better model.

If AIC had not been used in this paper, it could have been argued that the MLP might outperform the statistical model not because the neural model represented a more accurate model, but because the statistical model contained only three parameters, whereas the MLP contained 10 parameters (connection weights).

III. RESULTS

Seeds germinated from 10.3ºC to 38.2ºC with respect to both cultivars. On the whole, the patterns of distribution of germination rate values across the range of temperatures were similar for both cultivars (Figure 1). In addition, it was observed that the highest germination rate values were in the 30-35ºC temperature range for both cultivars.

Figure 1

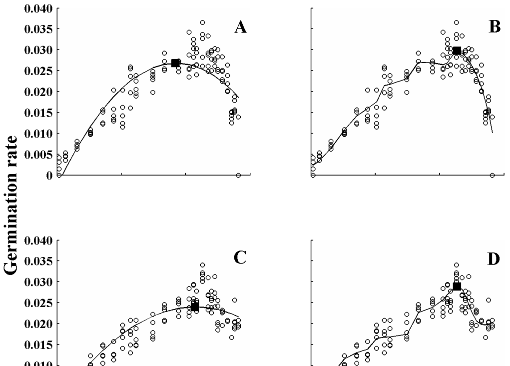

Figure Caption - Germination rate (in hours^{-1}) of two *Phaseolus vulgaris* cultivars as a function of 25 distinct temperatures. Circles represent the raw data set (n=125) for both cultivars (Carioca, **A&B**, and Iapar 81 **C&D**). The thin solid lines represent the predicted values derived from either the quadratic regression (**A&C**) or the multi-layer perceptron (**B&D**). The solid squares represent the optimum germination rate value derived either from using the quadratic regression (**A&C**) or from implementing Particle Swarm Optimization procedures (**B&D**).

The coefficients of determination (R^2) and root mean square error values (RMSE) derived from the data subsets utilized in both training and test procedures are available in Table 1. It was noted that the MLP neural network was more accurate than the quadratic regression in relation to both cultivars because it provided higher R^2 and lower RMSE values, as well as lower AIC values than the statistical model regarding both the training and test subsets.

The Iapar 81 data set analyzed by means of the MLP exhibited the highest R^2 and the lowest root mean square error values with respect to both training (R^2=0.908 and RMSE=0.0022) and test (R^2=0.901 and RMSE=0.0023) subsets, as well as the lowest AIC value for the test data set (Table 1). On the other hand, the Carioca data set modeled by means of the statistical regression produced the least accurate performance regarding both training (R^2=0.771 and RMSE=0.0041) and test (R^2=0.684 and RMSE=0.0052) subsets, as well as the highest AIC values.

Table 2 shows the optimum germination rate as well as their related temperature values derived from implementing two distinct methods (PSO and quadratic regression). With respect to both cultivars, Figure 1 reveals that the optimum germination rate values derived from using the statistical model (A&C) were less accurate than those obtained by means of the PSO method (B&D). The optimum germination rate values derived from the quadratic regression and the actual (observed) maximum values do not match (A&C), especially in relation to the Carioca data set (A). On the other hand, the optimum germination rate values derived from using the PSO method were more realistic regarding both cultivars (B&D).

Table 1. Accuracies

Algorithm	Metric	Carioca	Iapar 81
MLP	**R^2**	0.905/0.878	0.908/0.901
	AIC	-546.82/-356.30	-572.83/-390.09
	RMSE	0.0026/0.0033	0.0022/0.0023
Quadratic	**R^2**	0.771/0.684	0.795/0.84
	AIC	-494.32/-324.03	-524.64/-382.51
	RMSE	0.0041/0.0052	0.0034/0.0029

Coefficients of determination (R^2), root mean square errors (RMSE), and AIC (Akaike's Information Criterion) values derived from implementing MLP (Multi-layer perceptron neural network) and quadratic regression for prediction of germination rate. 'Carioca' and 'Iapar 81' are bean cultivars. Results are shown as training result/test accuracies.

Table 2 Optimum germination rate and temperature derived from implementing PSO and quadratic regression.

		Carioca	Iapar 81
PSO	GR_o	0.030	0.029
	T_o	32.8 ºC	32.7 ºC
Quadratic Regression	GR_o	0.027	0.024
	T_o	28.4 ºC	31.4 ºC

GR_o: optimum germination rate.
T_o: optimum temperature.
PSO: Particle Swarm Optimization.
Quadratic regression: polynomial regression, a general linear statistical model.
'Carioca' and 'Iapar 81' are bean cultivars.

IV. DISCUSSION AND CONCLUSIONS

Most biological processes in plants might well be characterized as intricate phenomena [9], and thus complex models such as ANNs and PSO could outperform traditional statistical methods. In fact, neural network techniques have been effectively used in modeling complex non-linear interactions between biological and ecological variables [24, 37].

For instance, [9] assessed the usefulness of a MLP, a relatively simple neural model, in a study of the effects of light intensity and sucrose concentration on the proliferation of kiwi microshoots. The outcome of the MLP was compared with the results derived from three statistical models (Poisson regression, logistic regression, and analysis of variance), and the neural network model was deemed to be an effective analytical tool. In the current paper, a MLP outperformed a conventional statistical model (Table 1). In addition, other agrobiological studies have demonstrated the superiority of MLP over traditional statistical procedures. For example, [22] utilized a MLP neural network for predicting tomato moisture ratio during its drying process, and the ANN was more accurate than empirical correlations.

Linear models were used in assessing relationships between germination speed parameters and seed traits of weed seeds. However, such linear models left a considerable amount of the total variance unexplained [10], since the R^2 values were lower than 0.6. It is a well-known fact that analyzing the germination of weed seeds in a thorough manner is essential for understating the biology and ecology of weed species [10, 31]. Therefore, neural modeling techniques could have been used in modeling and analyzing such germination parameters and traits in order to try to explain a larger proportion of the total variance. In the current work, the neural model (MLP) explained a large proportion of the total variance, since the R^2 values for the test subsets were higher than 0.87. On the other hand, the R^2 values derived from the quadratic regression (a linear model), using the same test examples as those used in implementing the neural model, were relatively low, especially for Carioca seeds (Table 1).

In this investigation, both cultivars showed similar patterns of distribution of germination rate values (Figure 1), and the PSO was able to highlight this fact in an accurate manner, since the optimum germination value for Carioca seeds were virtually the same as the optimum value for Iapar 81 (Table 2). On the other hand, the quadratic regression was not able to reveal such a fact, especially in relation to Carioca seeds (Table 2, Figure 1-A).

In order to determine the optimum range of germination rate values, [4] and [17] implemented a large number of non-parametric statistical tests (Mann-Whitney test) and used the outcomes derived from such tests as if they constituted one single statistical test. However, this should be deemed to be incorrect from a statistical perspective, because the probability of obtaining a significant difference when, in reality, the samples being compared come from the sample population is too high [2, 35]. Hence, from a statistical standpoint, the predictive models utilised in the present investigation constitute more effective, realistic procedures than those utilised by [4] and [17].

It should be noted that [4] stated that the optimum germination rate value for the Carioca beans should be in the 30.5-34.5ºC temperature range, and the optimum germination rate for Iapar 81 should be 32.5ºC. Table 2 shows that the optimum germination values derived from the neural modeling techniques are similar to those derived from the non-parametric statistical procedure implemented by [4]. Nonetheless, their outcomes were based on an inaccurate statistical procedure.

It is important to highlight that neural models can outperform statistical methods in analyzing both complex phenomena [24] and simpler phenomena, as was the case in this investigation.

It could be argued that a quadratic regression may not be able to fit the germination rate of seeds in an accurate manner. Figure 1 (A&C) shows that the maximum germination rate values are not where the quadratic function says (Table 2). However, a MLP is a simple ANN model, and thus a simple neural model outperformed a classical statistical method in this paper. Therefore, the current investigation provides statistical evidence of the ineffectiveness of standard statistical models in modeling the germination rate of seeds, and we recommend that researchers should deploy both more complex neural modeling techniques and higher-order regression models in order to analyze the germination of crop seeds.

REFERENCES

[1] V. Alvarado and K. J. Bradford. A hydrothermal time model explains the cardinal temperatures for seed germination. *Plant, Cell and Environment* 25, 1061-1069, 2002.

[2] A. Bianconi, J. S. Govone, B. F. J. Manly and M. J. Watts. The use of a multivariate statistical procedure in analysing the germination process of two bean cultivars, compared with a univariate approach. *Proceedings of the International Academy of Ecology and Environmental Sciences* 1, 70-76, 2011.

[3] T. A. Bytnerowicz and R. I. Carruthers. Temperature-dependent models of *Zannichellia palustris* seed germination for application in

aquatic systems. *Environmental and Experimental Botany*, 2014 (In press).

[4] V. J. M. Cardoso and A. Bianconi. Hydrotime model can describe the response of common bean (*Phaseolus vulgaris* L.) seeds to temperature and reduced water potential. *Acta Scientarum – Biological Sciences* 35, 255-261, 2013.

[5] G. R. Chantre, A. M. Blanco, M. V. Lodovichi, A. J. Bandoni, M. R. Sabbatini, R. L. López, M. R. Vigna and R. Gigón. Modeling *Avena fátua* seedling emergence dynamics: An artificial neural network approach. *Computers and Electronics in Agriculture* 88, 95-102, 2012.

[6] K. Donohue. Seeds and seasons: interpreting germination timing in the field. *Seed Science Research* 15, 175-187, 2005.

[7] B. P. Dubey, S. G. Bhagwat, S. P. Shouche and J. K. Sainis. Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosystems Engineering* 95, 61-67, 2006.

[8] F. Fernández-Luqueño, V. Reyes-Varela, C. Martínez-Suárez, G. Salomón-Hernández, J. Yáñez-Meneses, J. M. Ceballos-Ramírez and L. Dendooven. Effect of different nitrogen sources on plant characteristics and yield of common bean (*Phaseolus vulgaris* L.). *Bioresource Technology* 101, 396-403, 2010.

[9] J. Gago, L. Martínez-Núñez, M. Landín and P. P. Gallego. Artificial neural networks as an alternative to the traditional statistical methodology in plant research. *Journal of Plant Physiology* 167, 23-27, 2010.

[10] A. Gardarin, C. Dürr and N. Colbach. Prediction of germination rates of weed species: Relationships between germination speed parameters and species traits. *Ecological Modelling* 222, 626-636, 2011.

[11] P. H. Graham and P. Ranalli. Common bean (*Phaseolus vulgaris* L.). *Field Crops Research* 53, 131-146, 1997.

[12] R. J. Gummerson. The effect of constant temperatures and osmotic potentials on the germination of sugar beet. *Journal of Experimental Botany* 37, 729-741, 1986.

[13] F. Han, C. Zuo, W. Wu, J. Li and Z. Liu. Model Predictive Control of the Grain Drying Process. *Mathematical Problems in Engineering* 2012 (Article ID 584376), 1-12, 2012.

[14] M. Kashiri, A. D. Garmakhany and A. A. Dehghani. Modelling of sorghum soaking using artificial neural networks (MLP). *Quality Assurance and Safety of Crops & Foods* 4, 179-184, 2012.

[15] M. Keller and J. Kollmann. Effects of seed provenance on germination of herbs for agricultural compensation sites. *Agriculture, Ecosystems and Environment* 72, 87-99, 1999.

[16] M. H. Kutner, C. J. Nachtsheim and J. Neter. *Applied Linear Regression Models*, Fourth ed. McGraw-Hill/ Irwin, New York, 2004.

[17] L. G. Labouriau and M. Agudo. On the physiology of seed germination in *Salvia hispanica* L. I. Temperature effects. *Anais da Academia Brasileira de Ciências* 59, 37-56, 1987.

[18] X. Li, Y. He, and C. Wu. Non-destructive discrimination of paddy seeds of different storage age based on Vis/NIR spectroscopy. *Journal of Stored Products Research* 44, 264-268, 2008.

[19] K. Mavi, I. Demir and S. Matthews. Mean germination time estimates the relative emergence of seed lots of three cucurbit crops under stress conditions. *Seed Science & Technology* 38, 14-25, 2010.

[20] J. A. C. Michelangeli, M. Bhakta, S. A. Gezan, K. J. Boote and C. E. Vallejos. From flower to seed: Identifying phenological markers and reliable growth functions to model reproductive development in the common bean (*Phaseolus vulgaris* L.). *Plant, Cell & Environment* 36, 2046-2058, 2013.

[21] M. Mladenov and M. Dejanov. Application of neural networks for seed germination assessment. In: *9th WSEAS International Conference on Neural Networks (NN'08)*. Sofia, Bulgaria, pp. 67-72, 2008.

[22] K. Movagharnejad and M. Nikzad. Modeling of tomato drying using artificial neural network. *Computers and Electronics in Agriculture* 59, 78-85, 2007.

[23] N. Nouaouria, M. Boukadoum and R. Proulx. Particle Swarm Classification: A survey and positioning. *Pattern Recognition* 46, 2028-2044, 2013.

[24] J. M. Paruelo and F. Tomasel. Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. *Ecological Modelling* 98, 173-186, 1997.

[25] R. M. Pérez-Sánchez, E. Jurado, L. Chapa-Vargas and J. Flores. Seed germination of Southern Chihuahuan Desert plants in response to elevated temperatures. *Journal of Arid Environments* 75, 978-980, 2011.

[26] J. Pourreza and A. Bahrani. Estimating cardinal temperatures of milk thistle (*Silybum marianum*) seed germination. *American-Eurasian Journal of Agricultural & Environmental Sciences.* 12, 1030-1034, 2012.

[27] M. A. P. Ramalho and A. F. B. Abreu. Cultivares. In: Vieira, C., Paula Jr., T.J., Borém, A. (Eds.), Feijão. Editora UFV, Viçosa, pp. 415-436, 2006.

[28] M. A. Ranal and D. G. Santana. How and why to measure the germination process? *Revista Brasileira de Botânica* 29, 1-11, 2006.

[29] G. G. Roberto and G. Habermann. Morphological and physiological responses of the recalcitrant *Euterpe edulis* seeds to light, temperature and gibberellins. *Seed Science & Technology* 38, 367-378, 2010.

[30] S. S. Seefeldt, K. K. Kidwell and J.E. Waller. Base growth temperatures, germination rates and growth response of contemporary spring wheat (*Triticum aestivum* L.) cultivars from the US Pacific Northwest. *Field Crops Research* 75, 47-52, 2002.

[31] M. Singh, A. H. M. Ramirez, S. D. Sharma and A. J. Jhala. Factors affecting the germination of tall morning glory (*Ipomoea purpurea*). *Weed Science* 60, 64-68, 2012.

[32] P. Waligórski and M. Szaleniec. Prediction of white cabbage (*Brassica oleracea* var. *capitata*) self-incompatibility based on neural network and discriminant analysis of complex electrophoretic patterns. *Computational Biology and Chemistry* 34, 115-121, 2010.

[33] M. J. Watts and S. P. Worner. Using Artificial Neural Networks to Determine the Relative Contribution of Abiotic Factors Influencing the Establishment of Insect Pest Species. In: *Ecological Informatics* 3(1) 64-74, 2008.

[34] R. Webster and A. B. McBratney. On the Akaike Information Criterion for choosing models for variograms of soil properties. *Journal of Soil Science* 40, 493-496, 1989.

[35] J. H. Zar. *Biostatistical Analysis*, Fourth ed. Prentice Hall Inc., New Jersey, 1999.

[36] J. R. Zhang, J. Zhang, T. M. Lok and M. R. Lyu. A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training. *Applied Mathematics and Computation* 185, 1026-1037, 2007.

[37] W. J. Zhang. Computational Ecology: Artificial Neural Networks and Their Applications. World Scientific Publishing Co. Pte. Ltd., Singapore, 2010.