Autonomous Agent Response Learning by a Multi-Species Particle Swarm Optimization

Chi-kin Chow and Hung-tat Tsui
Visual Signal Processing and Communications Laboratory
Department of Electronic Engineering
The Chinese University of Hong Kong
E-mail: {ckchow1, httsui}@ee.cuhk.edu.hk

Abstract – **A novel autonomous agent response learning (AARL) algorithm is presented in this paper. We proposed to decompose the award function into a set of local award functions. By optimizing this objective function set, the response function with maximum award can be determined. To tackle the optimization problem, a modified Particle Swarm Optimization (PSO) called "Multi-Species PSO (MS-PSO)" is introduced by considering each objective function as a specie swarm. Two sets of experiments are provided to illustrate the performance of MS-PSO. The results show that it returns a more accurate response set within shorter duration by comparing with other PSO methods.**

I. INTRODUCTION

In robot applications such as factory automation and autonomous vehicles, the learning algorithms [1, 2] of autonomous agents gain more attentions in the recent years. An autonomous agent is able to self-improve its response behavior so as to adopt to the environment. The response behavior *R* can be represented as an vector function of observation from the environment: **p** = *R*(**o**) where **p** is the response vector and **o** is the observation vector. Most literatures [3-7] express the response function as a state diagram that each observation is defined as a state. The weights among the states are trained by some statistical learning methods such as Reinforcement Learning [3-6] and Hidden Markov Model [7]. This representation suffers from the problem that the observation states are insufficient to describe the continuous environment of the real-world applications. To tackle this problem, continuous representations of response function were suggested [8, 9], and neural network is a common approach [10]. By considering an observations **o** and its desired response **r** as training sample **B** = [**o** | **r**], the continuous response function can be established through training the neural network on sample set {$\mathbf{B}_i$}. In the view of autonomous agent, the desired response of an observation can be self-extracted from the target award function *A*(**o**, **r**) in which all extracted responses satisfy the following constraint:

$$\text{Response Constriant}: A(\vec{o}, R(\vec{o})) \geq A(\vec{o}, \vec{r}) \qquad \text{Eq. (1)}$$

In this paper, we model the response extraction process as a multiobjective optimization problem. By using the correlation property among the objective functions, a modified PSO called "Multi-Specie PSO (MS-PSO)" is proposed to speed up the searching process. Afterward, the response function is constructed by training the observation-response samples with Support Vector Regression [11].

The remaining of this paper is organized as follows: The architecture of the proposed AARL algorithm is presented at section 2. In section 3, a population-based optimization method "Particle Swarm Optimization (PSO)" and its variants are reviewed. A modified PSO called "Multi-Species PSO (MS-PSO)" is proposed for solving the multiobjective optimization problem in section 4. Two environments are provided for response learning in order to illustrate the performance of the proposed algorithm at section 5. And a conclusion is drawn in section 6.

II. THE PROPOSED RESPONSE LEARNING ALGORITHM

To deal with the continuous environment robot applications, a continuous response representation is necessary. In this paper, the response function is expressed as a Gaussian Mixture Model (GMM) *U*(**o**):

$$U(\vec{o}) = \sum_{i=1}^{N_u} M_i e^{-\frac{\|\vec{o}-\vec{\mu}_i\|}{2\sigma_i^2}}$$

where M_i, σ_i, $\boldsymbol{\mu}_i$ are the weight, variance and mean of the i^{th} Gaussian function. Based on the response constraint described at Eq. (1), the response learning process can be formulated as an optimization problem:

$$\max_{\{\vec{W}_i\}} \int A(\vec{o}, U(\vec{o})) \partial \vec{o} \qquad \text{Eq. (2)}$$

where $\vec{W}_i = [M_i, \sigma_i, \vec{\mu}_i]$

Since Eq. (2) is extremely complicated, we suggest an alternative approach to construct the response network by a set of observation-response (O-R) samples $\{\mathbf{B_i}\} = [\mathbf{o_i} \mid \mathbf{r_i}]$ in which $\mathbf{o_i}$ and $\mathbf{r_i}$ satisfy the response constraint Eq. (1). In order to provide a good sample set for constructing the response network, the samples should be uniformly distributed on the observation domain. By fixing the observation as $\mathbf{o}$, the award function can be rewritten as an Local Award Function (LAF) $A_o(\mathbf{r}) = A(\mathbf{o}, \mathbf{r})$. Hence, the award function is decomposed to the LAF set based on the observations of O-R samples. By optimizing the LAF set, the response of O-R samples can be determined. In addition, since the award function is assumed to be smooth and continuous, $\mathbf{r_j}$ is close to $\mathbf{r_i}$ if the distance between $\mathbf{o_i}$ and $\mathbf{o_j}$ is small. Therefore, the response learning algorithm can be formulated as an multi-objective optimization problem in which the optima are correlated. After extracting the response of the O-R samples, response network is determined by training the samples. The procedure of the proposed AARL algorithm is summarized as follows:

Step 1: Define the size of O-R sample set (N_s).
Step 2: Define the observation vector $\mathbf{o_i}$ and hence the local award function $A_i(\mathbf{r})$ of $\mathbf{B_i}$.
Step 3: Search the optimal $\{\mathbf{r_i}\}$ of the LAF set by the proposed MS-PSO.
Step 4: Construct the response network by training the O-R samples.

III. REVIEWS OF PARTICLE SWARM OPTIMIZATIONS

A. Conventional Particle Swarm Optimization

Particle Swarm Optimization [12] (PSO) algorithm is based on the metaphor of individuals refining their knowledge by interacting with one another. A particle is a moving point in an n-dimensional solution space. Besides its position $\mathbf{x_i}$ and velocity $\mathbf{v_i}$, each particle stores the best position in the search space it has found thus far in a velocity $\mathbf{v_i}$. The velocity of the particle is adjusted stochastically toward its previous best position and the best position found by any member of its neighborhood:

$$\mathbf{v_i} \leftarrow \mathbf{v_i} + \alpha_1(\mathbf{p_i} - \mathbf{x_i}) + \alpha_2(\mathbf{p_g} - \mathbf{x_i}) \qquad \text{Eq. (3)}$$

where α_1 and α_2 sensitivities to the local best position and global best position respectively. The index g is the index of the particle in the neighborhood with the best performance so far. Hence $\mathbf{p_g}$ is the best vector found by the swarm. Once $\mathbf{v_i}$ has been calculated, the particle's position $\mathbf{x_i}$ is adjusted as:

$$\mathbf{x_i} \leftarrow \mathbf{x_i} + \mathbf{v_i} \qquad \text{Eq. (4)}$$

The algorithm is often compared to the family of evolutionary computations, as a stochastic population-based search of a problem space. The PSO differs from evolutionary methods in an important way, however: it does not implement selection of the fittest. Instead, individuals persist over time, and adapt by changing. To handle the multiobjective optimization at AARL, the objective function is formulated as the linear combination of the LAF set:

$$F(\mathbf{x}) = A_1(\mathbf{r_1}) + A_2(\mathbf{r_2}) + \ldots + A_n(\mathbf{r_n})$$

where the particle is represented as $\mathbf{x} = [\mathbf{r_1}\ \mathbf{r_2}\ \ldots\ \mathbf{r_n}]$. The disadvantage of this approach is that it is sensitive to the domination of some LAFs.

B. MultiObjective PSO

Due to the drawback of the conventional PSO on optimizing a multiobjective function, some researchers [13, 14] proposed the modified PSOs to handle this problem. Carlos [14] proposed the "Multi-Objective Particle Swarm Optimization (MO-PSO)" by introduceing the Pareto ranking scheme [15]. In this algorithm, the historical record of best solutions found by a particle could be used to store non-dominated solutions generated in the past. The use of global attraction mechanisms combined with the historical archive of previously found non-dominated vectors would motivate convergence towards globally non-dominated solutions. Based on the idea of having a global repository, every particle will deposit its flight experiences after each flight cycle. In addition, the updates to the repository are performed considering a geographically based system defined in terms of objective function values of each particle.

IV. MULTI-SPECIES PARTICLE SWARM OPTIMIZATION

In this section, we describe the details of Multi-Species Particle Swarm optimization (MS-PSO). It takes the advantage of MO-PSO that a non-dominated solution is returned. In addition, by using the correlation among the objective functions, a shorter computational time is needed. Different from the conventional PSO that all particles aim at optimizing an objective function, the particles of MS-PSO optimize multiple functions concurrently. In the MS-PSO model, each objective function forms a specie, which is a subpopulation that optimizes an specific objective function. Therefore, the number of swarm specie equals to the size of objective function set. In addition, a communication channel is established between the neighbor swarms for transmitting the information of best particles, in order to provide guidance for improving their objective values. MSPSO involves 3 types of attraction from:

i. Its historical local best position.
ii. Global best particle of its current specie.

iii. Global best particle of its neighbor species.

where the third attraction is newly introduced in PSO. The velocity of jth particle in the swarm $\mathbf{S}_i$ is adjusted as follows:

$$\mathbf{v}_{i,j} \leftarrow \mathbf{v}_{i,j} + \alpha_1(\mathbf{p}_{i,j} - \mathbf{x}_{i,j}) + \alpha_2(\mathbf{p}_{i,g} - \mathbf{x}_{i,j}) + \alpha_3\mathbf{n}_i \quad \text{Eq. (5)}$$

where $\mathbf{n}_i$ is the Neighbor Swarm Reference Velocity (NSRV) of swarm $\mathbf{S}_i$:

$$\vec{n}_i = \sum_{k=1}^{H_i}(\vec{a}_{i,k} - \vec{x}_{i,j})$$

H_i is the neighbor size of $\mathbf{S}_i$, $\mathbf{a}_{i,k}$ is the best particle of k^{th} neighbor swarm of $\mathbf{S}_i$ and α_3 is the Neighbor Influence constant. Figure 1 shows an example of a MS-PSO model. This model consists of 4 different species $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$, which aim at searching the optimum of the objective functions A_1, A_2 A_3 and A_4 respectively. Moreover, there are 4 communication channels among the swarms $\mathbf{S}_1$ - $\mathbf{S}_2$, $\mathbf{S}_1$ - $\mathbf{S}_3$, $\mathbf{S}_1$ - $\mathbf{S}_4$ and $\mathbf{S}_2$ - $\mathbf{S}_4$.

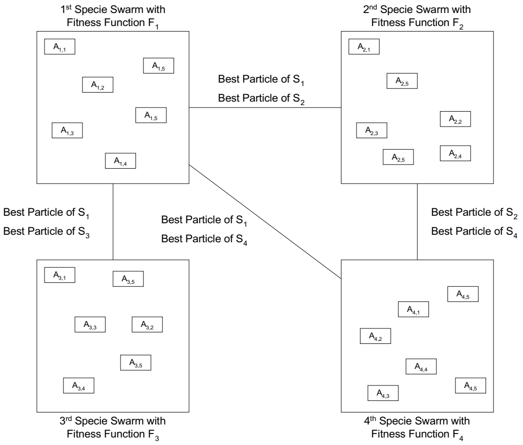


Figure 1 An Example of MSPSO

The procedure of finding the response in AARL with N_s O-R samples by the MS-PSO is summarized as follows:

Step 1: Construct a MS-PSO model with N_s species.

Step 2: Define the neighbor(s) of the swarms. The swarm $\mathbf{S}_j$ is regarded as the neighbor swarm of $\mathbf{S}_i$ if their observation distance is smaller than the Observation Influence constant η

$$\text{i.e. } |\mathbf{o}_i - \mathbf{o}_j| < \eta.$$

Step 3: Randomly initialize N_p particles with uniform distribution in each swarm.

Step 4: The particles keep interacting with others inside the same swarm and adjusting its positions until the change of improvement approximately equals to zero.

V. EXPERIMENTAL RESULTS

In this section, we apply the proposed AARL algorithm to extract the desired responses of two environments. The performance of the proposed algorithm is concluded by 2 measurements. The first measurement is the performance of MS-PSO on finding the responses of the OR samples, which is defined as:

$$P_S \approx \sum_{i=1}^{N_o}(U(\vec{o}_i) - R(\vec{o}_i))^2 \qquad \text{Eq. (6)}$$

where N_o is the O-R sample size. The second measurement is called Response Network Performance (RNP). By given any observation $\mathbf{o}$, the response network $U(\mathbf{o})$ should return a response with maximum amount of award. Therefore, the RNP can be evaluated from its sum of award P_r:

$$P_r = \int A(\vec{o}, U(\vec{o}))d\vec{o} \qquad \text{Eq. (7)}$$

To speed up the process on measuring the RNP, the Eq. (7) is approximated as:

$$P_r \approx \begin{cases} \sum_{i=0}^{N_d}(U(i\Delta o) - R(i\Delta o))^2 & \text{1D case} \\ \sum_{i=0}^{N_d}\sum_{j=0}^{N_d}(U(i\Delta o, j\Delta o) - R(i\Delta o, j\Delta o))^2 & \text{2D case} \end{cases} \qquad \text{Eq. (8)}$$

where N_d is the number measurement division and Δo is the measurement division width. In order to evaluate the noise sensitivity of the proposed algorithm, the experiments will be repeated by introducing a zero-mean Gaussian noise with magnitude 0.1 and variance 0.1 into the environments. For further, we repeat the experiments by applying the conventional PSO and MO-PSO, in order to illustrate the contributions of MS-PSO. To provide a fair comparison, the population size of conventional PSO and MO-PSO are set as the total population of MS-PSO. The parameters $[\alpha_1\ \alpha_2\ \alpha_3]$ are set as 0.1. In the performance measurement, N_d and Δo are assigned as 41 and 0.025 respectively. All simulations are process at the PC platform with 1.7GHz CPU with 256MB memory.

A. 1-Observation and 1-Response Award Function

In this experiment, a 1-Observation, 1-Response award function was learnt by the proposed algorithm:

$$A_1(o,r) = 0.5\cos(d)\exp\left(-\frac{d^2}{8\pi^2}\right) \qquad \text{Eq. (9)}$$

where $d = 0.5(\sin(2\pi o) + 1) - r$ and $o, r \in [0, 1]$. The corresponding desired response function is:

$$R_1(o) = 0.5(\sin(2\pi o) + 1)$$

Figure 2 shows the surface plot of the target award function in which the point with higher intensity indicates its relative larger award value. It is observed that the local award functions formed by Eq. (9) consist of multiple optima. The observation sample sets of size 5, 10, 15 and 20 are generated uniformly on the observation domain, and hence the MSPSO with 5, 10, 15, 20 species are constructed. Moreover, we assigned 5 particles in each specie swarm for searching its optima. Fig. 3 illustrates an example of observation sample distribution with 20 species and 5 particles in each swarm. After searching the responses of the O-R samples, the corresponding response networks are construct by training the sample sets with Support Vector Regression. Figure 4a shows the resultant response network outputs. As the size of observation sample set increases, a more accurate response network can be achieved. Table 1 and 3 listed the comparisons on the speed and accuracy among the PSOs. The results show that the proposed MS-PSO is the fastest approach among them. In additional, it can return the optima with lower objective value than the others.

B. 2-Observation and 1-Response Award Function

In this experiment, the performance of the proposed algorithm on learning a 2-Observation and 1-Response award function was studied.

$$A_2(o_1,o_2,r) = 0.75\cos(d)\exp\left(-\frac{d^2}{2\pi^2}\right) \quad \text{Eq. (10)}$$

where $d = \pi\,(sin(2\pi o_1) + 1)(cos(2\pi o_2) + 1) - 4\pi r$ and o_1, o_2, $r \in [0, 1]$. The corresponding desired Response functions are:

$$R_2(o_1,o_2) = 0.25(\sin(2\pi o_1)+1)(\cos(2\pi o_2)+1)$$

Fig. 5 shows the award values lie on the 4 planes: o_1 = 0.5, o_2 = 0.2, o_2 = 0.8 and r = 0.8. 4 sets of observation sample with sizes 49, 100, 196 and 289 are provided for extracting the corresponding responses. Table 2 and 4 listed the comparisons on the accuracy and the speed among the PSOs. It is found that the computational time of MS-PSOs is much less than that of others methods while the accuracy is as good as MO-PSO. Figure 7 shows the resultant response network output of the observation sample sizes equal to 100 and 289. Similar to the pervious section, the experiment sets are repeated by adding the noise described in previous experiment fig. 8 shows the outputs of resultant response network.

VI. CONCLUSION

We presented a novel response learning algorithm on the parameterized environment in this article. The proposed algorithm is mainly divided into 3 steps: (1) Award Function Decomposition (2) Desired Response Extraction and (3) Response Network construction. In the decomposition stage, the given award function is decomposed by a set of predefined observation vectors. In order to form the training set for constructing the response function, the responses of the training set are determined by a newly proposed PSO call "Multi-Specie PSO (MS-PSO)". Afterward, the response function is constructed based on the samples extracted from the pervious 2 steps. In order to verify the proposed algorithm, two environments are evolved. The results show that the proposed MS-PSO returns a more accurate optima by using less computation time than that of the conventional PSO and MO-PSO.

REFERENCES

[1] Bentivegna, D.C.; Atkeson, C.G.; "Learning from observation using primitives", Proceedings of the IEEE International Conference on Robotics and Automation 2001, ICRA 2001, Volume: 2, 2001, Pages: 1988 – 1993.

[2] Enokida, S.; Ohashi, T.; Yoshida, T.; Ejima, T.; "Stochastic field model for autonomous robot learning", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics 1999, IEEE SMC '99, Volume: 2, 12-15 Oct. 1999, Pages: 752 – 757.

[3] Yamaguchi, T.; Masubuchi, M.; Fujihara, K.; Yachida, M.; "Realtime reinforcement learning for a real robot in the real environment", Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems '96, IROS 96, Volume: 3, 4-8 Nov. 1996, Pages: 1321 – 1328.

[4] Pack Kaelbling, L.; "On reinforcement learning for robots", Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems '96, IROS 96, Volume: 3, 4-8 Nov. 1996, Pages: 1319 – 1320.

[5] Howell, M.N.; Gordon, T.J.; "Associative reinforcement learning for discrete-time optimal control", Learning Systems for Control (Ref. No. 2000/069), IEE Seminar, 26 May 2000, Pages:1/1 - 1/4.

[6] Kaitwanidvilai, S.; Parnichkun, M.; "Active Bayesian feature weighting in reinforcement learning robot", 2002 IEEE International Conference on Industrial Technology, 2002, IEEE ICIT '02, Volume: 2, 11-14 Dec. 2002, Pages: 1090 – 1095.

[7] Ogawara, K.; Takamatsu, J.; Kimura, H.; Ikeuchi, K.; "Modeling manipulation interactions by hidden Markov model", IEEE/RSJ International Conference on Intelligent Robots and System, 2002, Volume: 2, 30 Sept.-5 Oct. 2002, Pages: 1096 – 1101.

[8] Schaal, S.; Atkeson, C.G.; "Robot learning by nonparametric regression", Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems '94, Advanced Robotic Systems and the Real World', IROS '94, Volume: 1, 12-16 Sept. 1994, Pages: 478 – 485.

[9] Atkeson, C.G.; "Using locally weighted regression for robot learning", Proceedings of the IEEE International Conference on Robotics and Automation 1991, ICRA 1991, 9-11 April 1991, Pages: 958 - 963.

[10] Gross, H.-M.; Stephan, V.; Krabbes, M.; "A neural field approach to topological reinforcement learning in continuous action spaces", IEEE World Congress on Computational Intelligence. The 1998 IEEE International

Joint Conference on Neural Networks Proceedings, 1998, Volume: 3, 4-9 May 1998, Pages: 1992 – 1997.

[11] V. N. Vapnik and C. Cortes, “Support Vector Networks”, Machine Learning, 20(3), 273-297, 1995.

[12] J. Kennedy and R. Eberhart; “Swarm Intelligence”, Morgan Kaufmann Publishers, 2001.

[13] Coello Coello, C.A.; Lechuga, M.S.; “MOPSO: a proposal for multiple objective particle swarm optimization”, CEC '02. Proceedings of the 2002 Congress on Evolutionary Computation, 2002, Volume: 2, 12-17 May 2002, Pages1051 – 1056

[14] Xiaohui Hu; Eberhart, R.; “Multiobjective optimization using dynamic neighborhood particle swarm optimization”, Proceedings of the 2002 Congress on Evolutionary Computation, 2002. CEC '02, Volume: 2, 12-17 May 2002, Pages: 1677 – 1681.

[15] David E. Goldberg; “Genetic Algorithms in Search, Optimization and Maching Learning”, Addison Wesley Publishing Company, Massachusetts, 1989.

TABLE 1

AVERAGE PERFORMANCE COMPARISON AMONG THE PSOS OF SECTION 5A

Number of Observation	**Conventional PSO**		**MO-PSO**		**MS-PSO**	
	Noise-Free	Noisy	Noise-Free	Noisy	Noise-Free	Noisy
5	0.0388	0.1902	10.01×10^{-4}	0.0053	5.69×10^{-4}	0.0049
10	0.1396	0.2840	11.01×10^{-4}	0.0072	8.92×10^{-4}	0.0085
15	0.1756	0.3905	25.01×10^{-4}	0.0196	20.0×10^{-4}	0.0142
20	0.2236	0.7824	29.01×10^{-4}	0.0284	23.4×10^{-4}	0.0210

TABLE 2

AVERAGE PERFORMANCE COMPARISON AMONG THE PSOS OF SECTION 5B

Number of Observation	**Conventional PSO**		**MO-PSO**		**MS-PSO**	
	Noise-Free	Noisy	Noise-Free	Noisy	Noise-Free	Noisy
49	3.805	4.408	0.0062	0.0143	0.0056	0.0178
100	5.98	7.854	0.0071	0.0372	0.0070	0.0325
196	8.45	21.044	0.0193	0.0873	0.0185	0.0774
289	12.66	24.793	0.0235	0.1536.	0.0248	0.1263

TABLE 3

AVERAGE COMPUTATIONAL TIME COMPARISON AMONG THE PSOS OF SECTION 5A

Number of Observation	**Conventional PSO**		**MO-PSO**		**MS-PSO**	
	Noise-Free	Noisy	Noise-Free	Noisy	Noise-Free	Noisy
5	3.51	8.78	4.44	8.47	1.45	3.08
10	8.58	26.29	23.71	33.57	2.43	5.68
15	18.39	50.55	42.47	78.36	6.58	10.02
20	28.26	87.07	71.28	145.86	6.88	11.12

TABLE 4

AVERAGE COMPUTATIONAL TIME (SEC.) COMPARISON AMONG THE PSOS OF SECTION 5B

Number of Observation	**Conventional PSO**		**MO-PSO**		**MS-PSO**	
	Noise-Free	Noisy	Noise-Free	Noisy	Noise-Free	Noisy
49	271.1	1489	198.9	1465	30.48	78.35
100	678.8	5003	623.3	5023	67.34	144.04
196	1616	15711	1669.4	15747	89.48	292.50
289	2663.3	30399	2660.2	30506	117.92	429.15

TABLE 5
AVERAGE RESPONSE NETWORK PERFORMANCE OF SECTION 5A

Number of Observation	Noise-Free	Noisy	Variance of SVR
5	0.2005	1.781	0.2
10	0.2221	1.687	0.15
15	0.1502	1.572	0.08
20	0.1432	1.472	0.05

TABLE 6
AVERAGE RESPONSE NETWORK PERFORMANCE OF AT SECTION 5B

Number of Observation	Noise-Free	Noisy	Variance of SVR
49	0.1915	0.7332	0.1
100	0.1572	0.5364	0.075
196	0.1169	0.6406	0.05
289	0.1430	0.5941	0.04

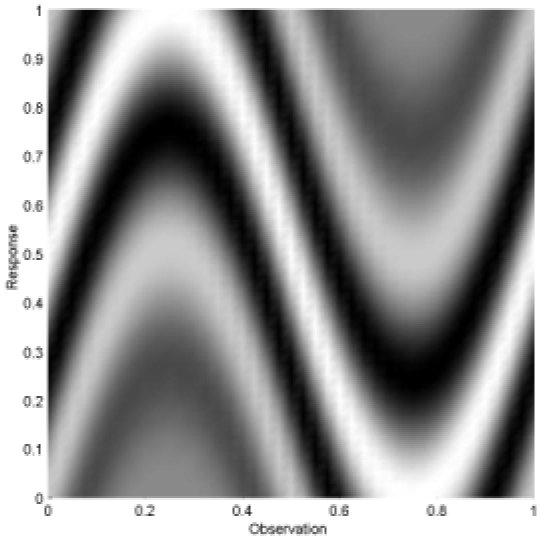


Figure 2 The award surfaces of section 5A

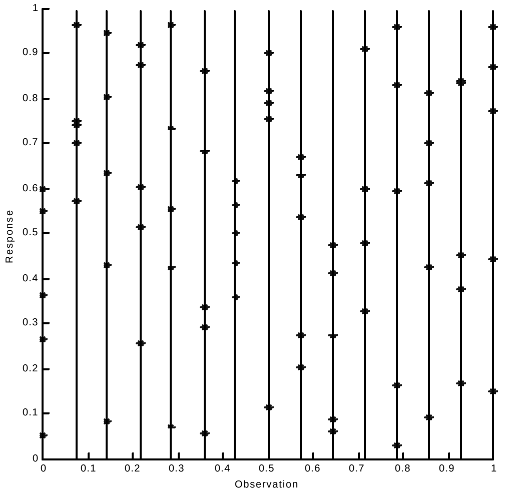
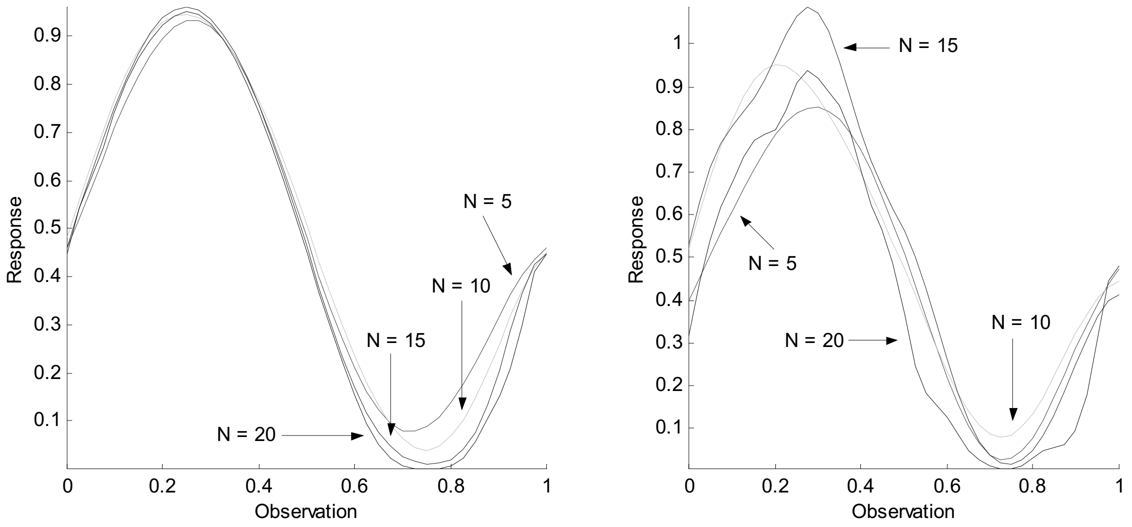


Figure 3 The initial particle distribution of swarms at section 5A



(a) (b)

Figure 4 Resultant Response Networks' Outputs (a) Noise-free observation samples (b) Noisy observation samples at section 5A

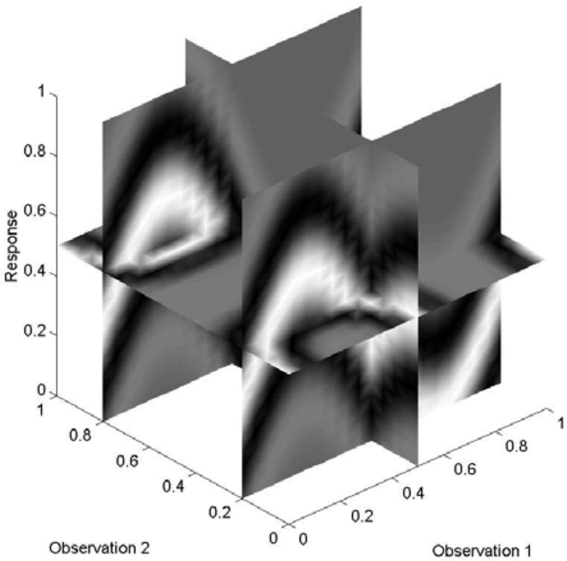


Figure 5 The slices of award volume at section 5B

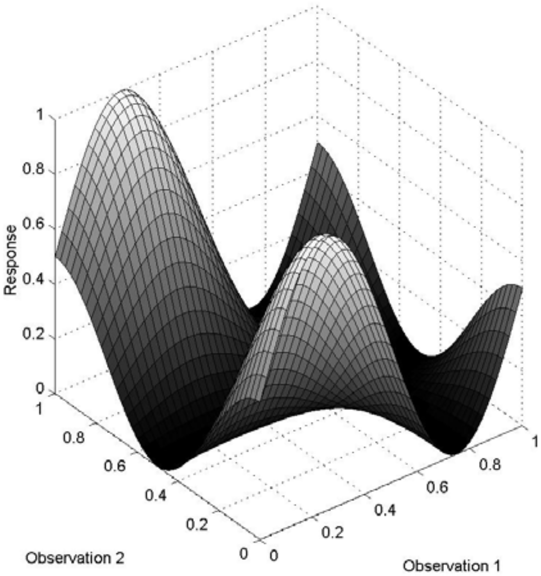
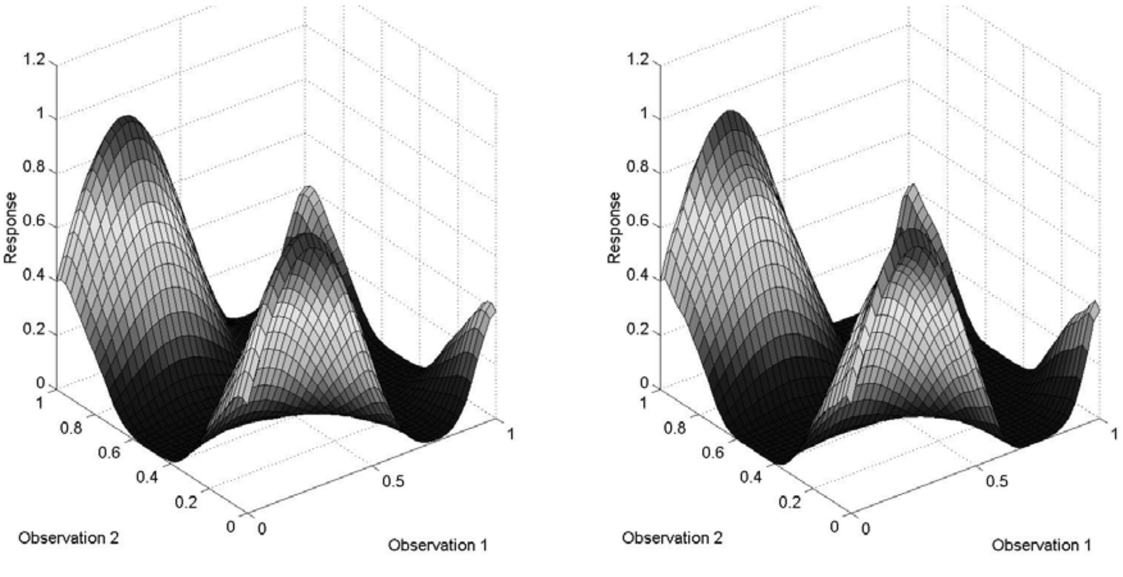
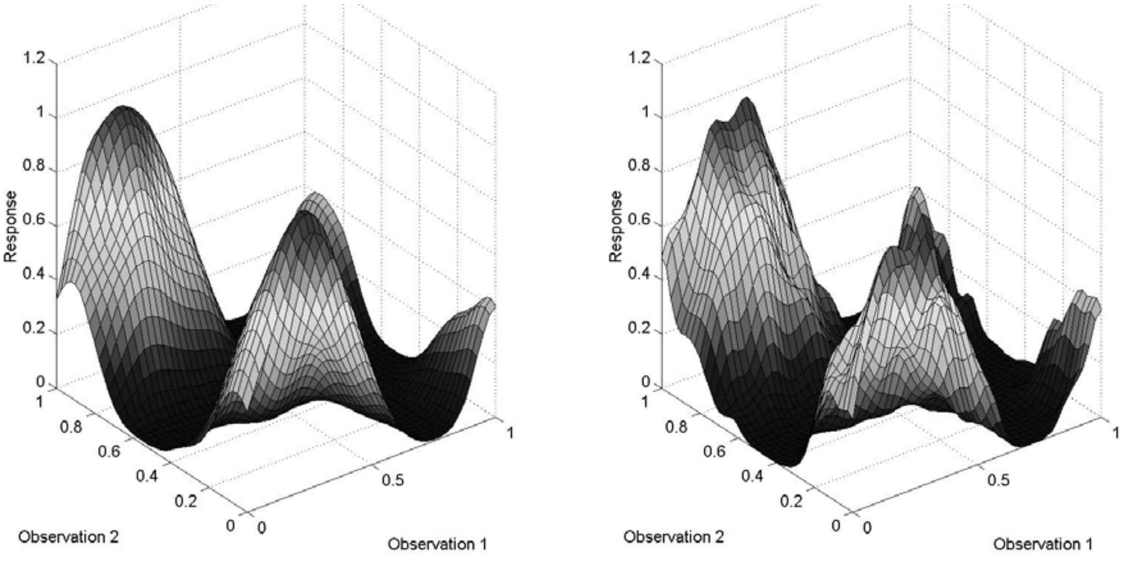


Figure 6 Desired Response Function of section 5B



(a) (b)

Figure 7 Outputs of Resultant Response Networks constructed by (a) 100 noise-free observation samples (b) 289 noise-free observation samples at Section 5B



(a) (b)

Figure 8 Outputs of Resultant Response Networks constructed by(a) 100 noisy observation samples (b) 289 noisy observation samples at Section 5B