

Human Opinion Dynamics Optimizer

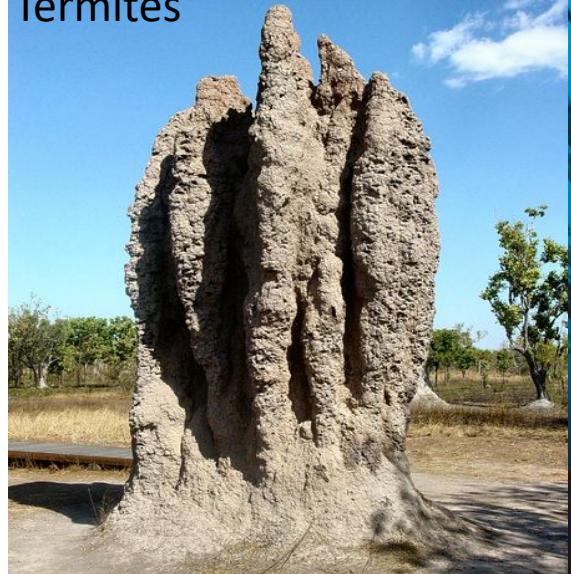
Rishemjit Kaur, PhD
Senior Scientist

CSIR-Central Scientific Instruments Organisation

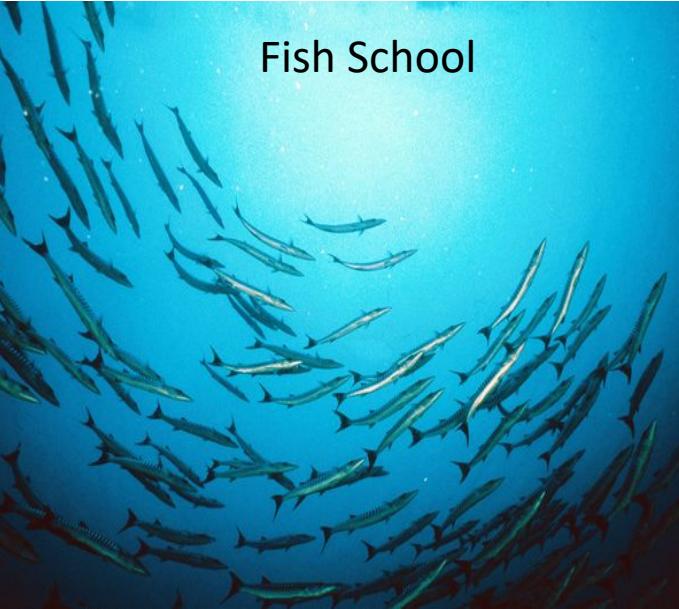
rishemjit.kaur@gmail.com, rishemjit.kaur@csio.res.in

Collective Intelligence in animals

Termites



Fish School



Flock of birds



Ant Colony



Animal herd



Collective Intelligence in Humans



Stock Market



Elections



Movie



Product



Restaurant

- Complex and intelligent animals on earth
- Governed by symbols or languages for communication
- Self-discourse
 - Organizes, gives sense to self and collective intelligence
- On a broader space-time

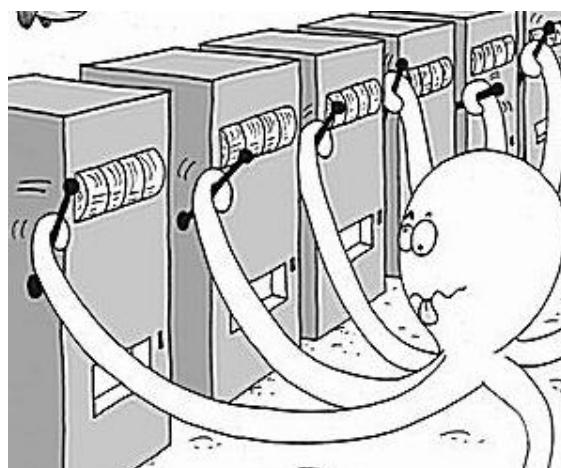
Human Collective Intelligence: Experimental studies



1907 Galton's experiment of collective intelligence[9]



Increased tax compliance rates
(Wenzel, 2005)[10]



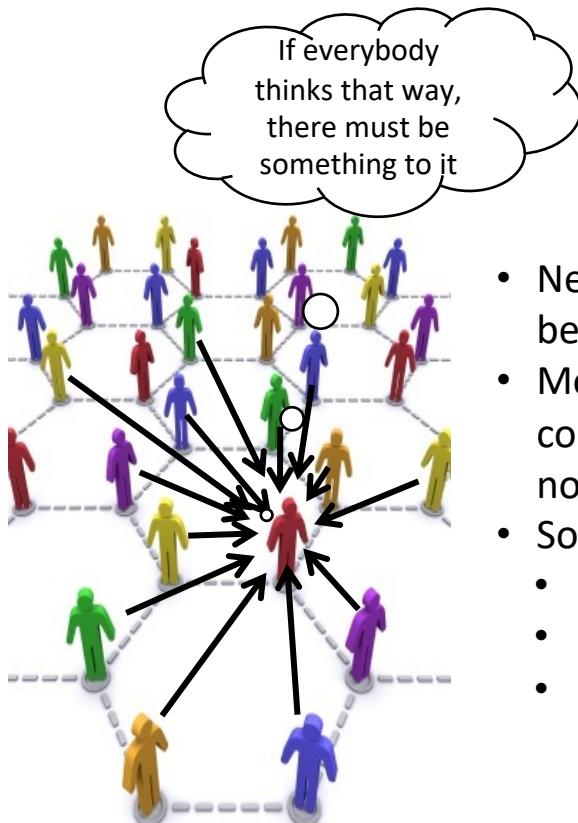
Multi armed Bandit problem (Toyokawa, 2014)[11]



Artificial "music market (Salganik,
2006)[12]

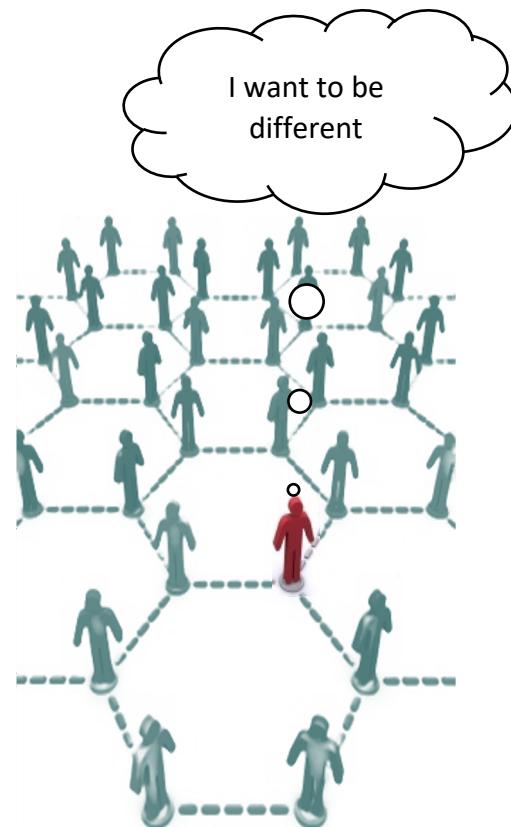
Continuous Opinion Dynamics Optimizer (CODO)

Durkheim's theory of social integration



- Need of belongingness
- Motivating to conform to norms
- Social Influence
 - Age
 - Immediacy
 - Number

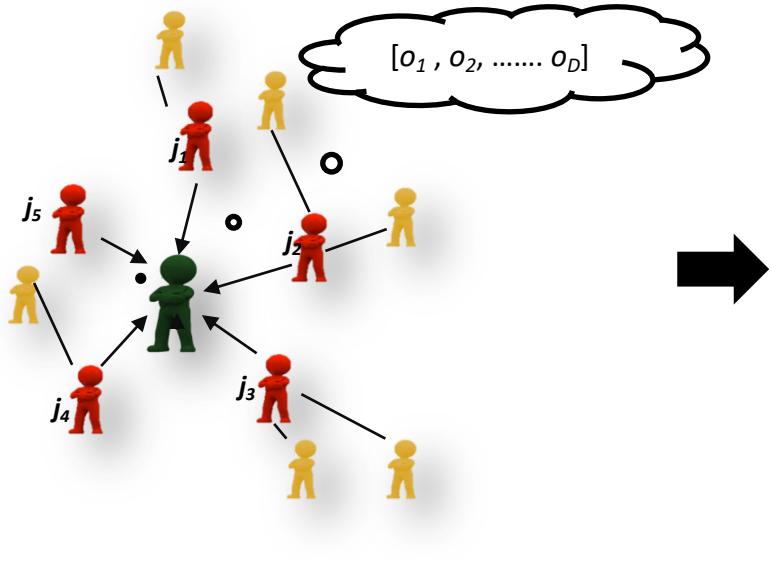
Integrating forces of social influence.



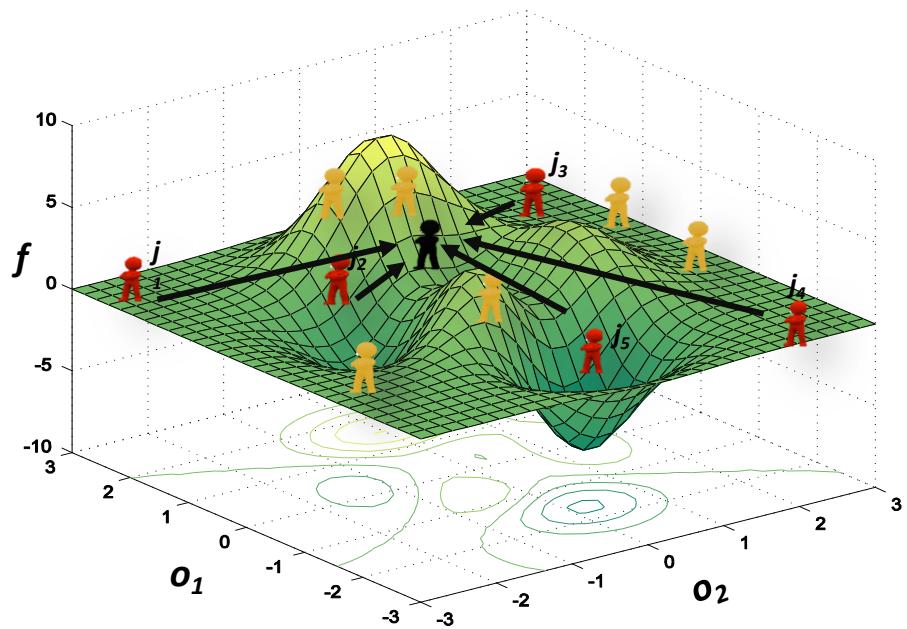
- Striving for Uniqueness
- Compensatory acts
- Consumer behaviour regarding fashions

Disintegrating forces of individualization

Continuous Opinion Dynamics Optimizer (CODO)

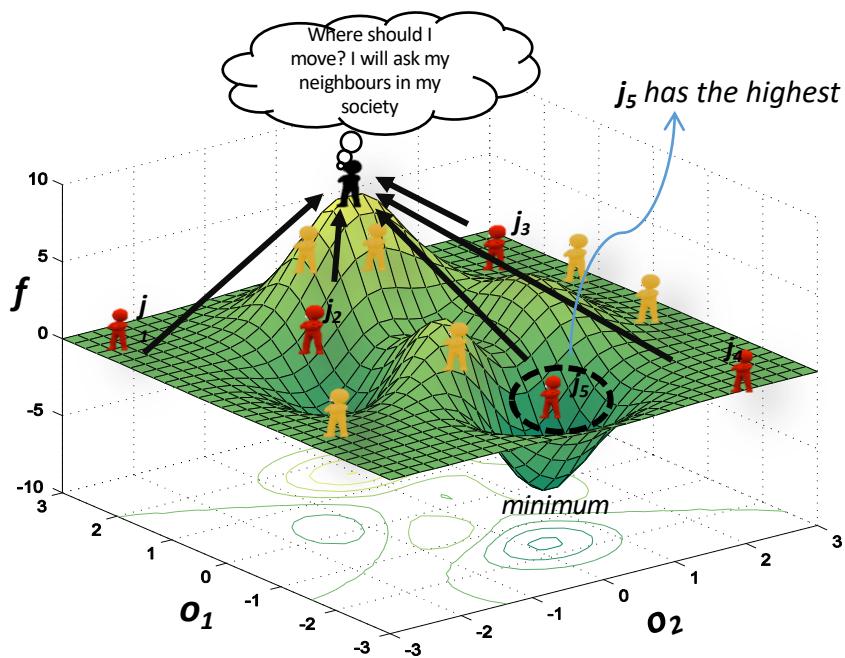


Societal Structure : Individuals j (red colour) are neighbours of individual i (black colour)

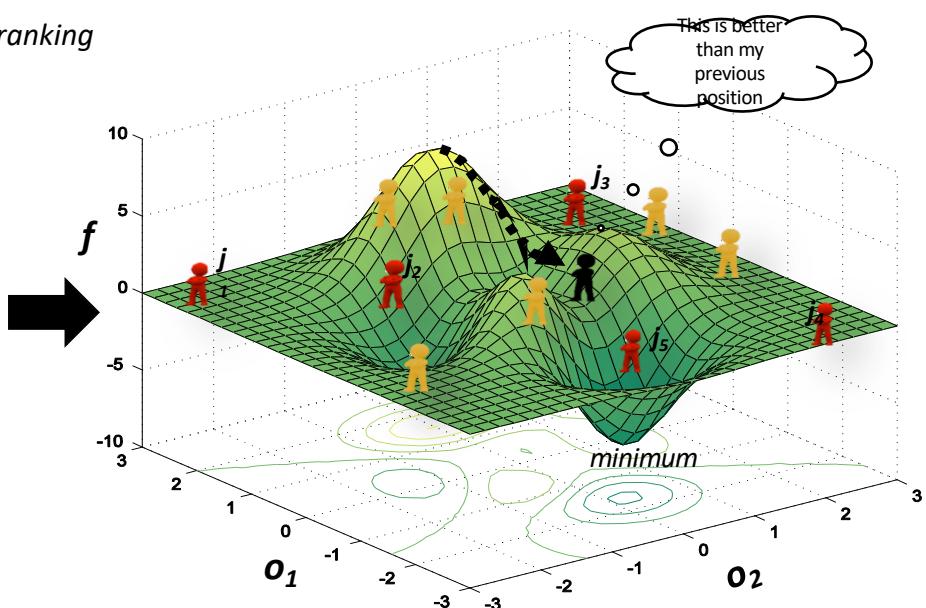


Mapping of individuals in search space/opinion space.

Integrating Forces



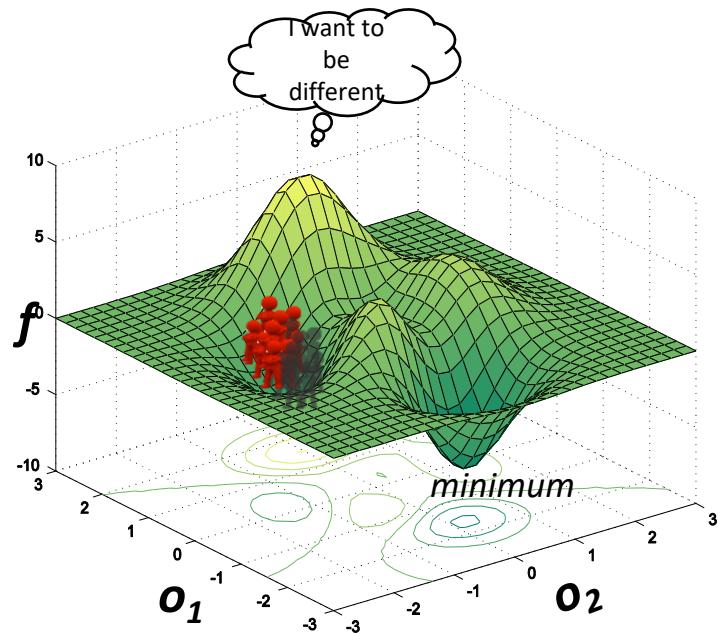
Initial positions of individuals in search space



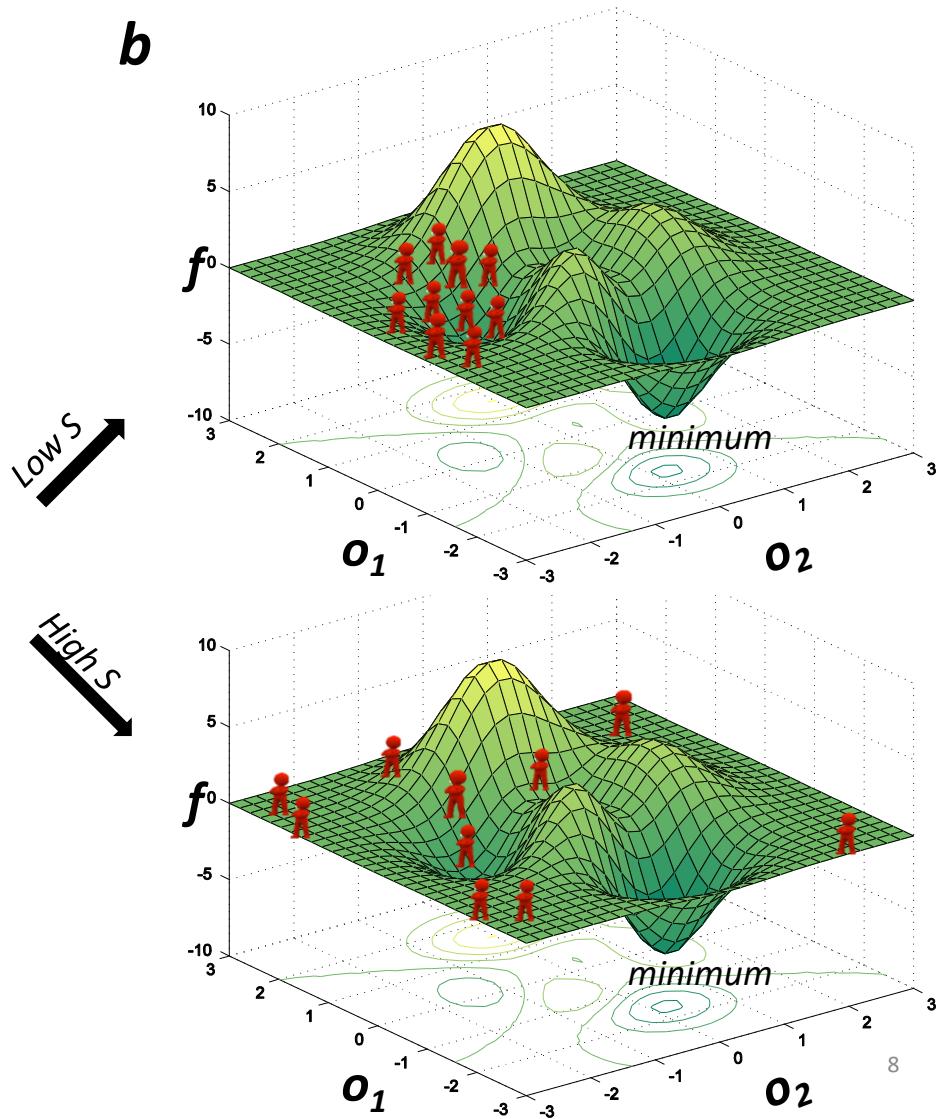
Updated position of individual i (black colour) after getting influenced by its neighbours.

Disintegrative forces

a



b

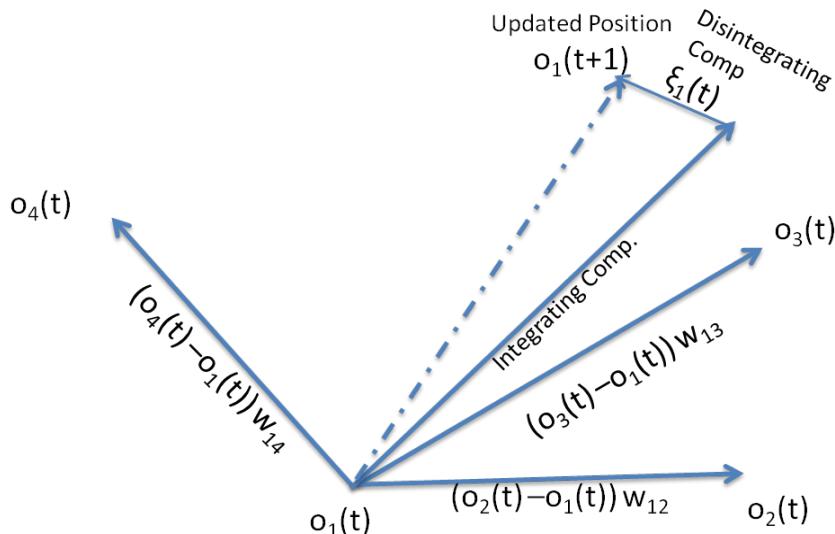


Updating Dynamics

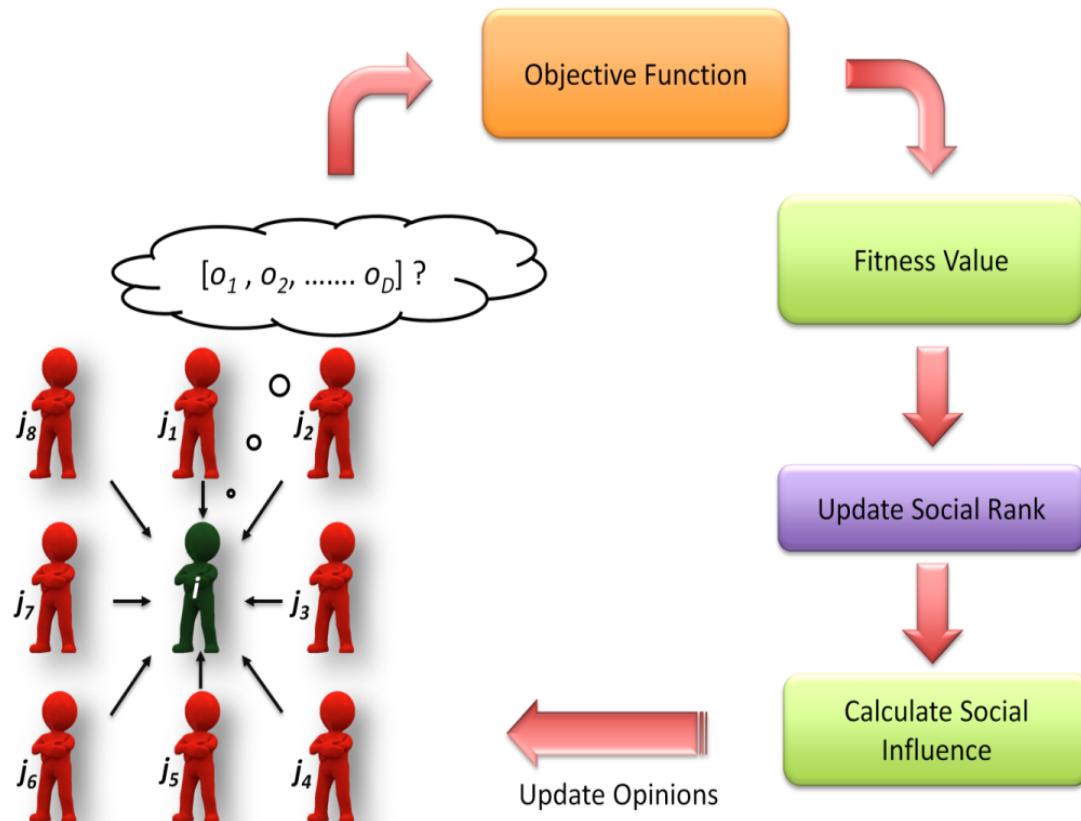
$$\Delta o_i = \frac{\sum_{j=1}^N (o_j(t) - o_i(t)) w_{ij}(t)}{\sum_{j=1}^N w_{ij}(t)} + \xi_i(t)$$

$w_{ij}(t) = \frac{SR_j(t)}{d_{ij}(t)}$ Social influence exerted by neighbour j on individual i
 $\xi_i(t) = N(0, \sigma_i(t))$ Distance between two individuals i and j
 $\sigma_i(t) = S \sum_{j=1}^N e^{-f_{ij}(t)}$ Adaptive noise with zero mean and std deviation
 $f_{ij} = |f_i - f_j|$ Strength of disintegrating forces

- **SR** Social ranking
 - is determined by their respective fitness values.
 - the highest **SR** is assigned to the individual with the minimum fitness value

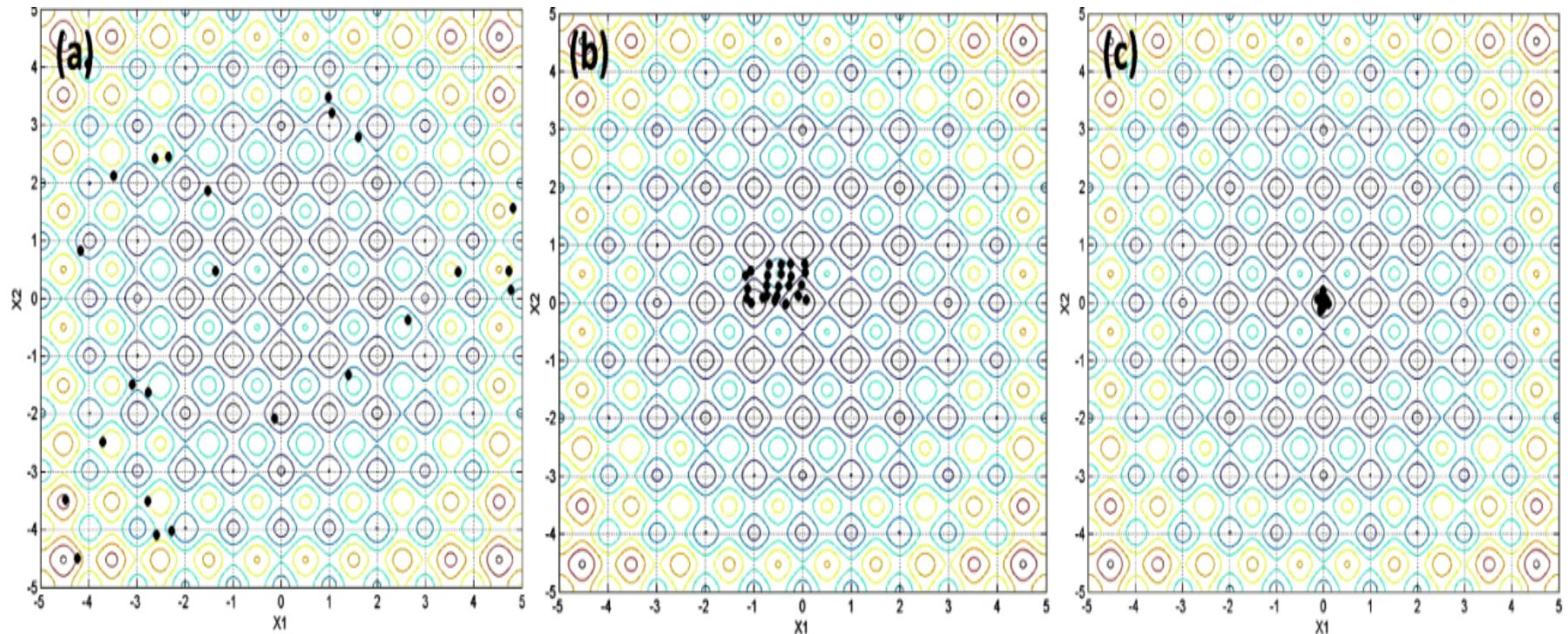


Optimization Process



Block diagram of optimization process

Working of CODO

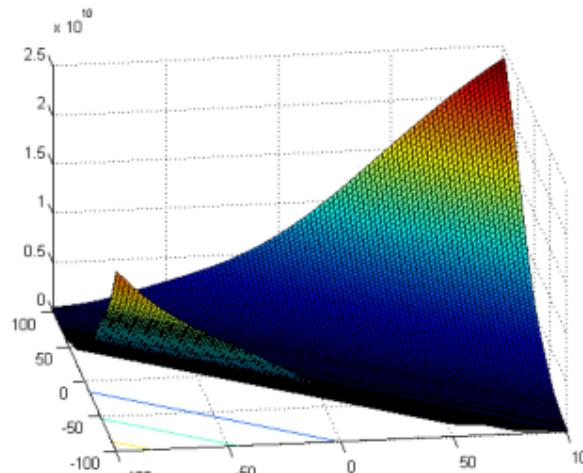


Population distribution during a) Individualization stage b) Exploration stage c) Exploitation stage at 25, 100 and 10000 function evaluations, respectively

Benchmark functions : CEC 2005 real-parameter

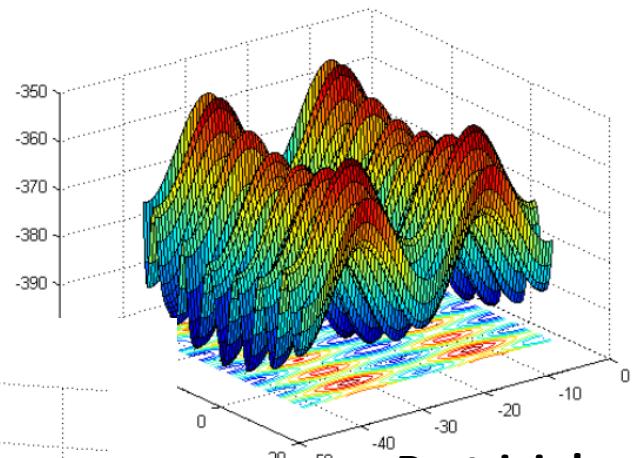
F	Name	Type	Separable	Initialization range	$f(x^*)$
F1	Spherical	U	S	$[-100, 100]^D$	-450
F2	Shifted Schwefel's Problem 1.2	U	NS	$[-100, 100]^D$	-450
F3	Shifted Rotated High Conditioned Elliptic	U	NS	$[-100, 100]^D$	-450
F4	Shifted Schwefel's Problem 1.2 with Noise in Fitness	U	NS	$[-100, 100]^D$	-450
F5	Schwefel's Problem 2.6 with Global Optimum on Bounds	U	NS	$[-100, 100]^D$	-310
F6	Shifted Rosenbrock's Function	BM	NS	$[-100, 100]^D$	390
F7	Shifted Rotated Griewank's Function without Bounds	BM	NS	$[0, 600]^D$	-180
F8	Shifted Rotated Ackley's Function with Global Optimum on Bounds	BM	NS	$[-32, 32]^D$	-140
F9	Shifted Rastrigin's Function	BM	S	$[-5, 5]^D$	-330
F10	Shifted Rotated Rastrigin's Function	BM	NS	$[-5, 5]^D$	-330
F11	Shifted Rotated Weierstrass Function	BM	NS	$[-0.5, 0.5]^D$	90
F12	Schwefel's Problem 2.13	BM	NS	$[-\pi, \pi]^D$	-460
F13	Shifted Expanded Extended Griewank's plus Rosenbrock's Function	EM	NS	$[-3, 1]^D$	-130
F14	Shifted Rotated Expanded Scaffer's F6	EM	NS	$[-100, 100]^D$	-300
F15	Hybrid Composition Function	HM	S near global optimum	$[-5, 5]^D$	120
F16	Rotated Hybrid Composition Function	HM	NS	$[-5, 5]^D$	120
F17	Rotated Hybrid Composition Function with Noise in Fitness	HM	NS	$[-5, 5]^D$	120
F18	Rotated Hybrid Composition Function	HM	NS	$[-5, 5]^D$	10
F19	Rotated Hybrid Composition Function with a Narrow Basin for the Global Optimum	HM	NS	$[-5, 5]^D$	10
F20	Rotated Hybrid Composition Function with the Global Optimum on the Bounds	HM	NS	$[-5, 5]^D$	10
F21	Rotated Hybrid Composition Function	HM	NS	$[-5, 5]^D$	360
F22	Rotated Hybrid Composition Function with High Condition Number Matrix	HM	NS	$[-5, 5]^D$	360
F23	Non-Continuous Rotated Hybrid Composition Function	HM	NS	$[-5, 5]^D$	360
F24	Rotated Hybrid Composition Function	HM	NS	$[-5, 5]^D$	260
F25	Rotated Hybrid Composition Function without Bounds	HM	NS	$[2, 5]^D$	260

Some examples of the functions



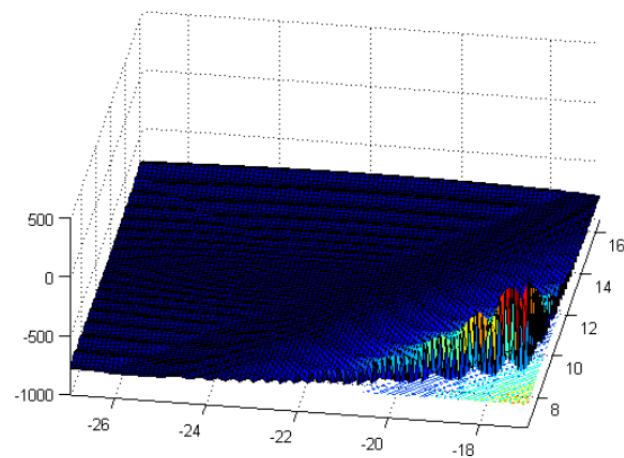
Rotated Discus

$$10^6 z_1^2 + \sum_{i=2}^D z_i^2 .$$



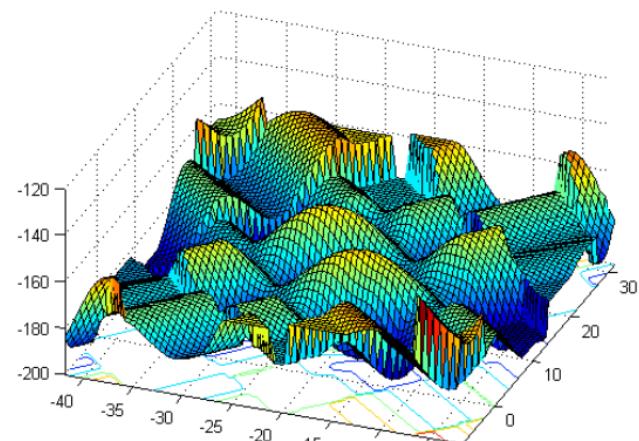
Rastrigin's

$$\sum_{i=1}^D (z_i^2 - 10 \cos(2\pi z_i) + 10) .$$

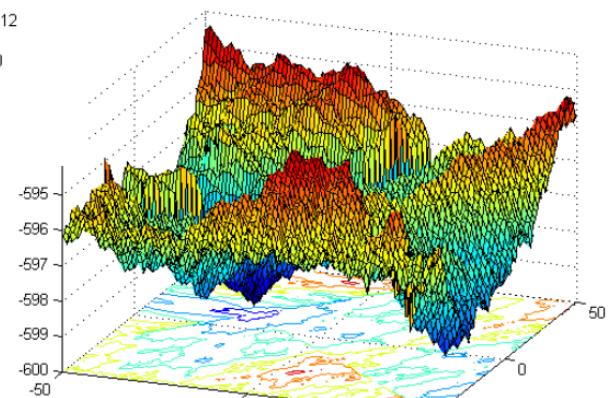


Rotated Schaffers

$$\left(\frac{1}{D-1} \sum_{i=1}^{D-1} (\sqrt{z_i} + \sqrt{z_i} \sin^2(50z_i^{0.2})) \right)^2 .$$



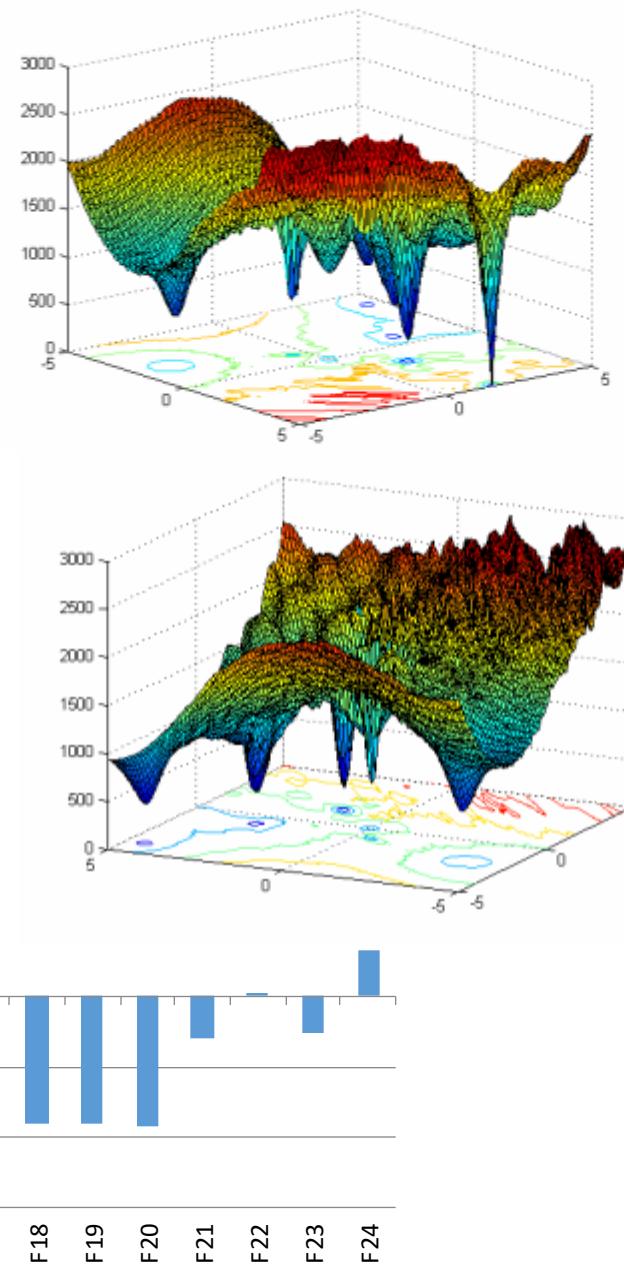
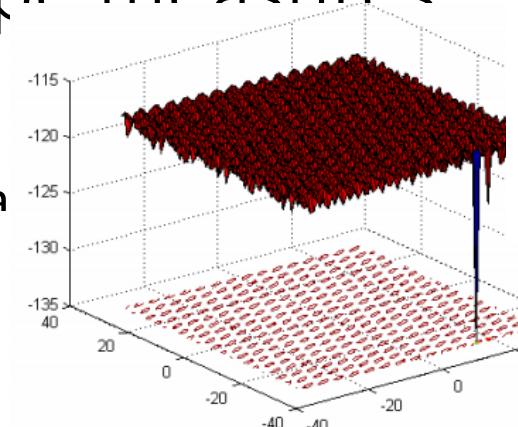
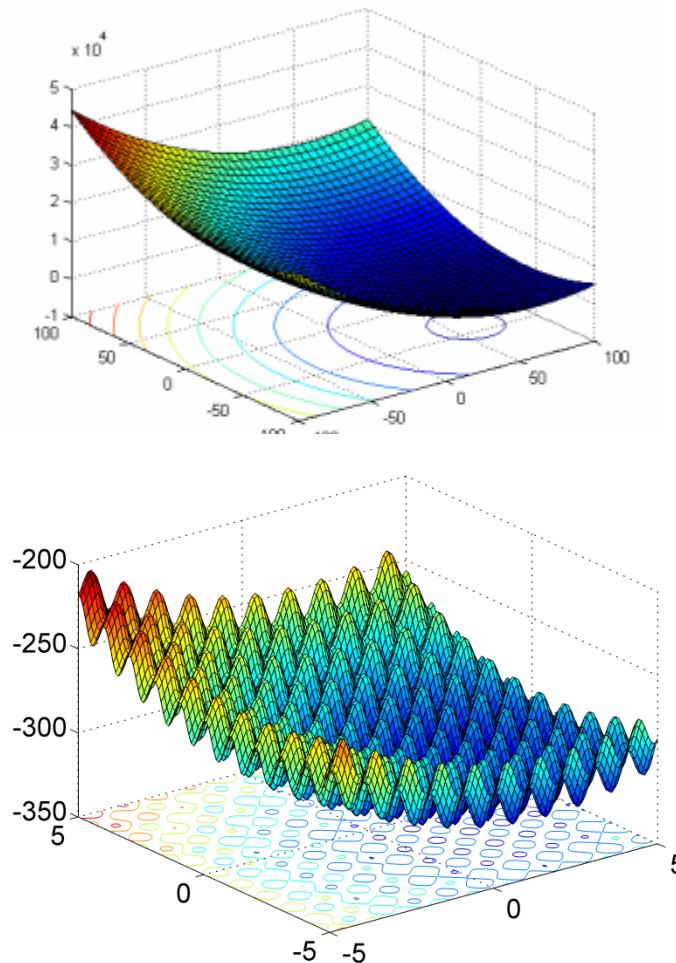
Non-continuous Rotated Rastrigin's



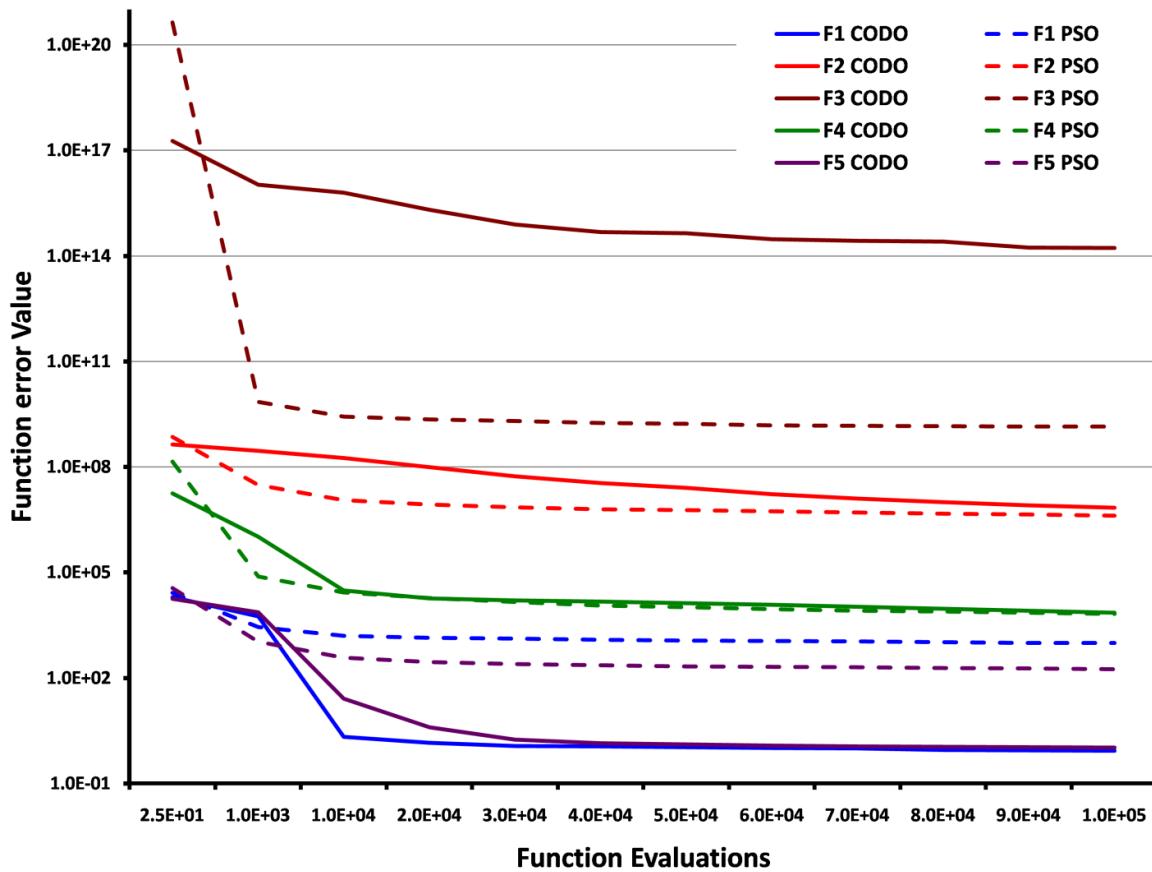
Rotated Weierstrass

Fitness Landscape measures

- Fitness distance correlation

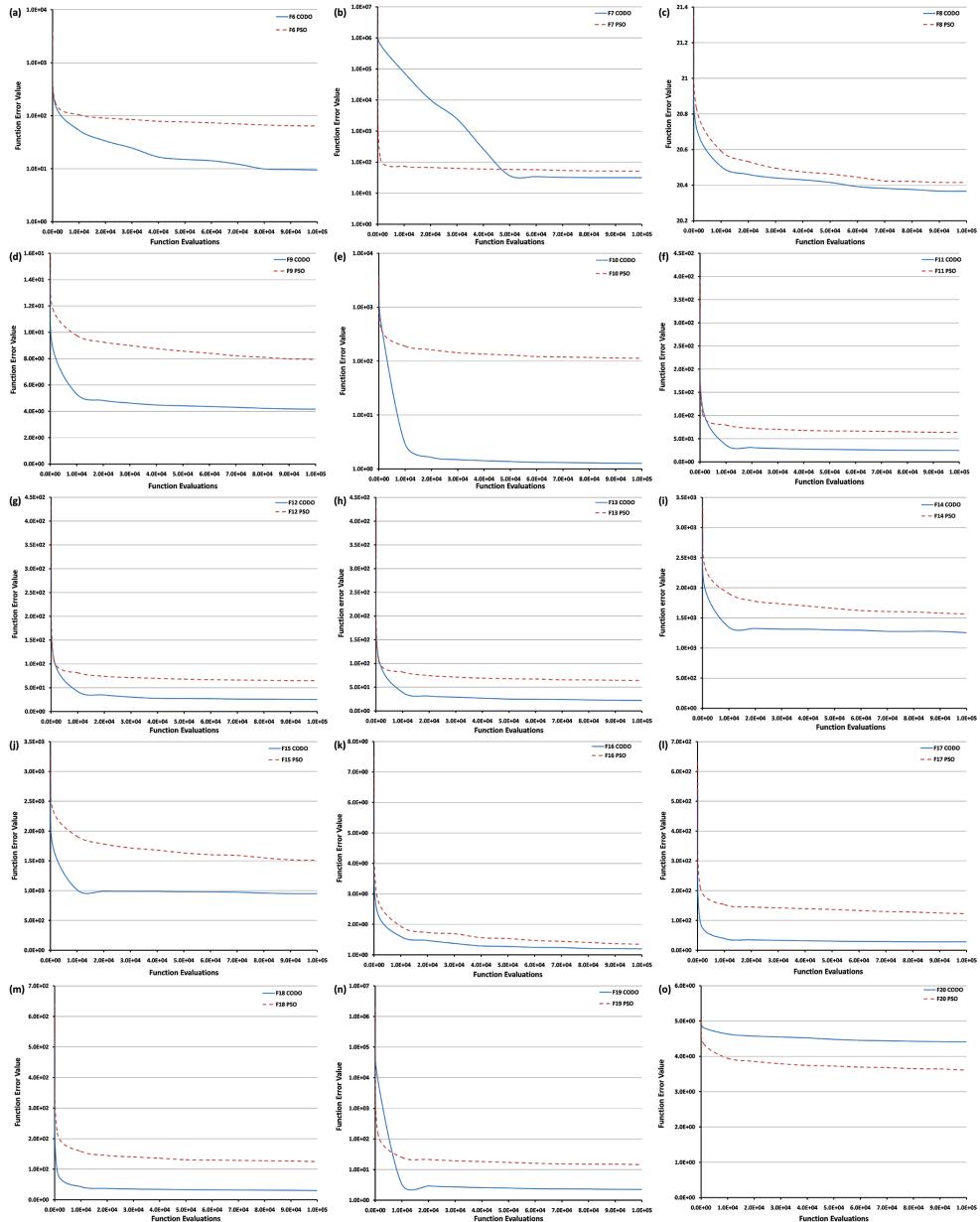


Unimodal Functions



- Relatively poor performance for F2 and F3 is observable due to the inherent characteristics of these functions
 - F2 is ill-conditioned function
 - F3 has a peculiar smooth but narrow ridge
- CODO performs better than lbest PSO in case of F1 and F5,
- Similar performance is obtained in case of F4.

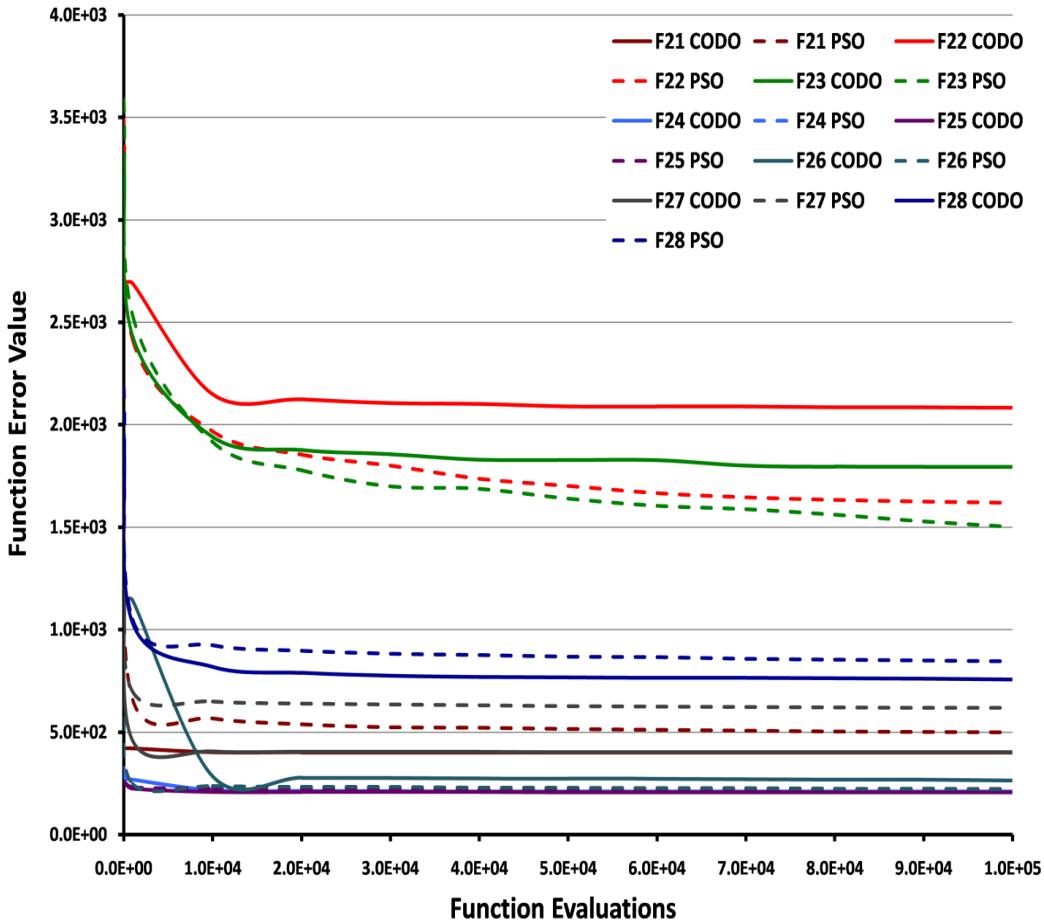
Multi-modal Functions



- Rotation invariance property of the CODO.

- F11 (Rastrigin's function) & F12 (Rotated Rastrigin's function),
- F14 (Schwefel's function) & F15 (Rotated Schwefel's function)
- F17 (Lunacek Bi_Rastrigin function) & F18 (Rotated Lunacek Bi_Rastrigin function)
- efficacy of algorithm in case of non continuous cases
- convergence results for F12 and F13 are similar in spite of F13 being a non-continuous version of F12.

Composition Functions



CODO and lbest PSO did not perform well on functions F22 and F23. This could be attributed to the fact that the basic function used in construction of these functions is Schwefel's function (F14) and rotated Schwefel's function (F15)

Effect of disintegrative forces

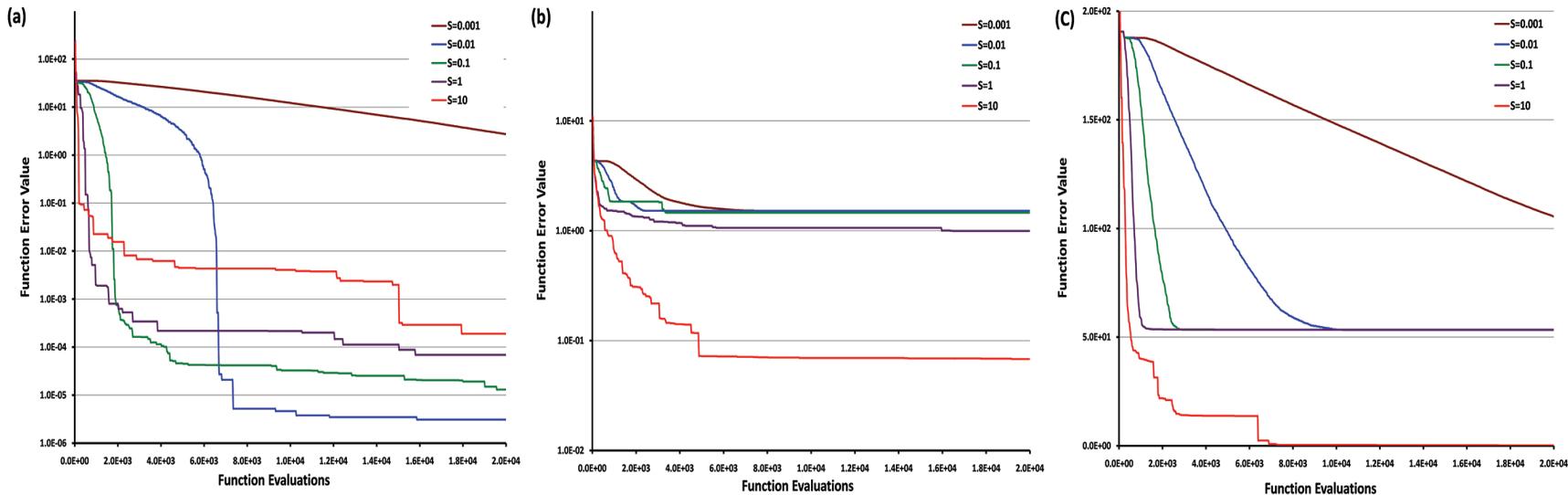


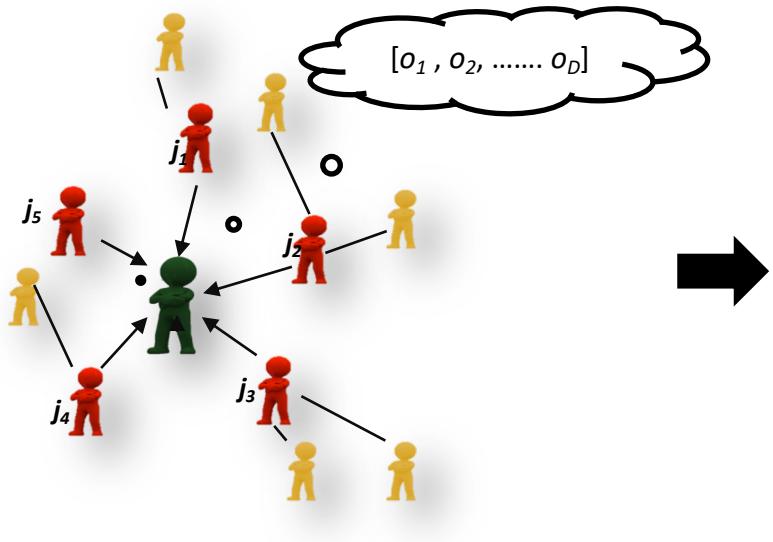
Figure (a, b and c) shows the convergence trends of functions from unimodal (F1), basic multimodal (F11) and composition (F21) categories

- Unimodal functions we get higher precision for the lower value of S .
- For multimodal functions (Fig. (b and c)), the larger the value of S , the better chances we have of reaching global minima. This may be attributed to the fact that larger S value brings with itself more diversity (greater individualisation) in the society and hence, the probability of getting stuck in local minima is less. Whereas, for lower values of S , the individuals tend to quickly gravitate towards consensus and get stuck in local minima.

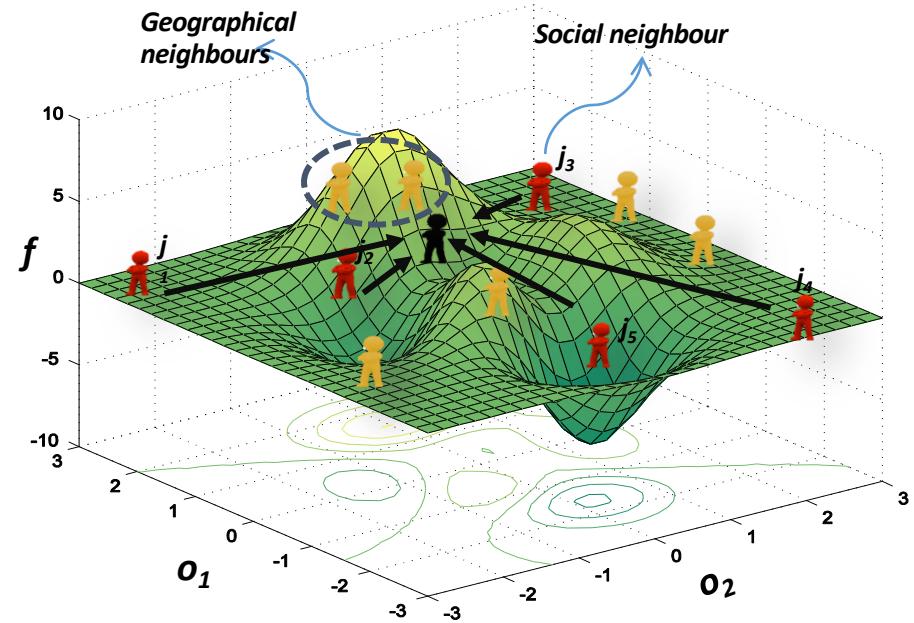
Open source library

- Open source library SitoLIB has been developed for human opinion formation based optimizers
- Available at <https://github.com/rishemjit/CODO>
- Two main functions:
 - `sitoOptimset`
 - `options = sitoOptimset('param1',value1,'param2',value2,...)`
 - `Sito`
 - `[x, fVal] = Sito(funct, nvars, options)`

Structure/ Topology



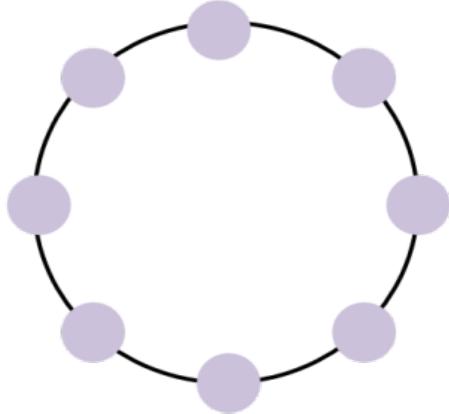
Societal Structure : Individuals j (red colour) are neighbours of individual i (black colour)



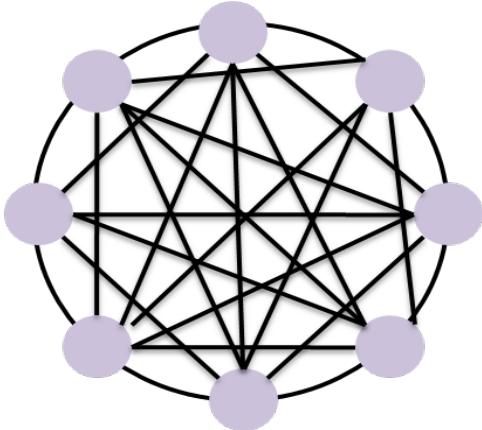
Mapping of individuals in search space/opinion space.

- Topology or structure affects *Convergence speed* as well as *Accuracy*
- Fully connected topology may result in early convergence and local optima.

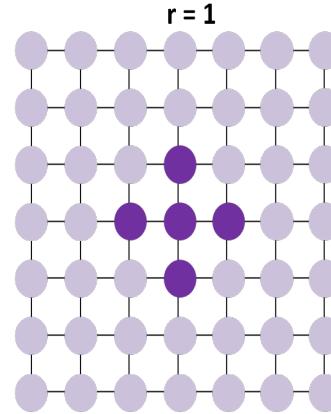
Types of topology



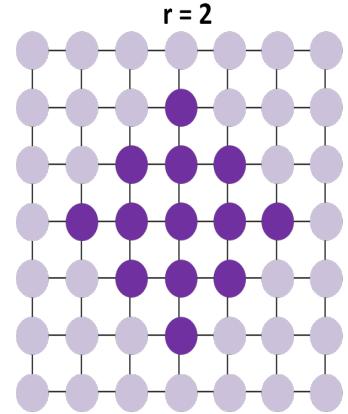
Ring



Fully-connected

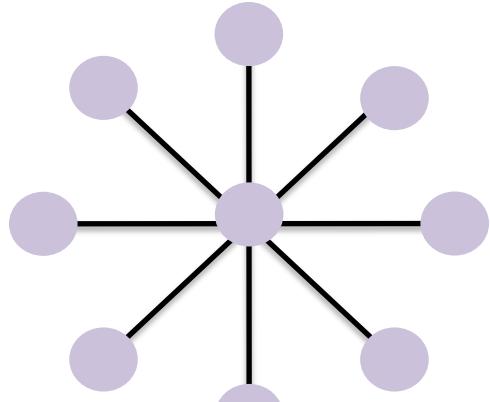


$r = 1$

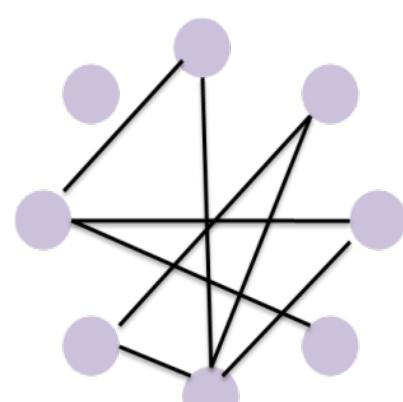


$r = 2$

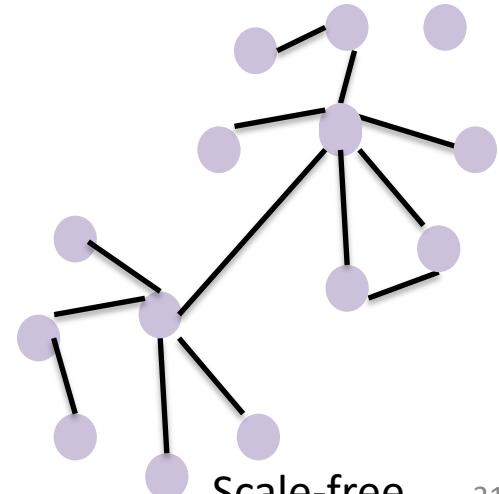
Mesh



Star

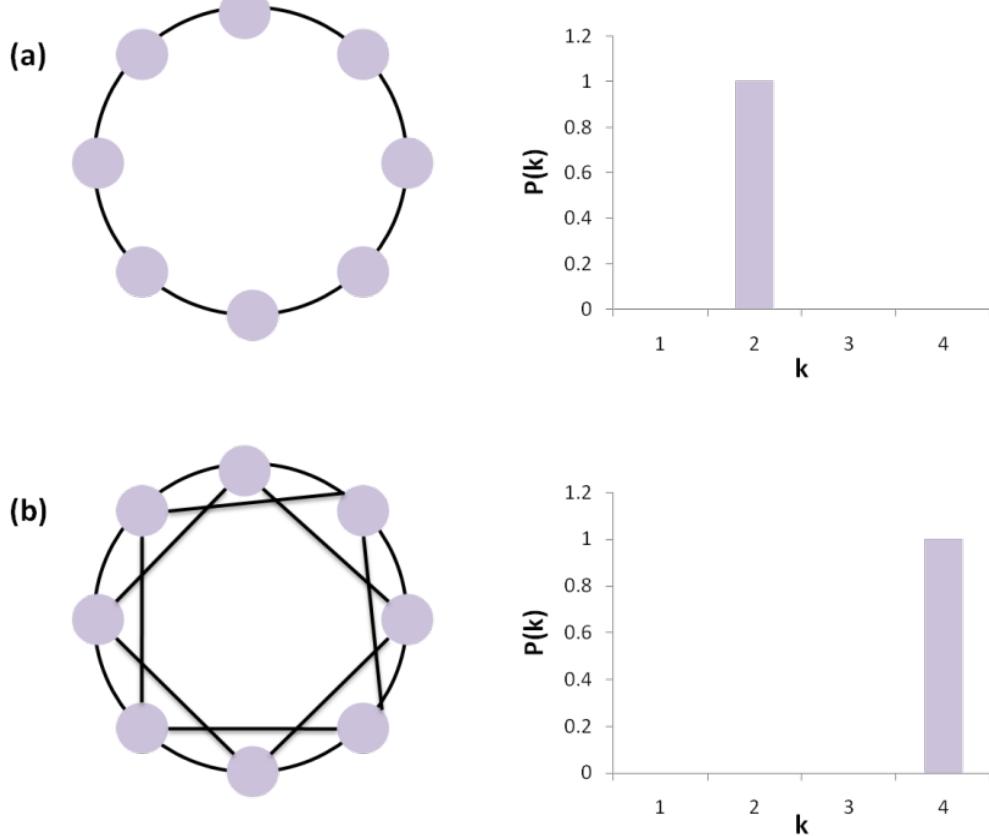


Random



Scale-free

Effect of topology size and neighborhood size



Society Size	Number of neighbors
10	2, 4, 6, 8, 10
20	2, 4, 6, 8, 10, 20
30	2, 4, 6, 8, 10, 20, 30
50	2, 4, 6, 8, 10, 20, 30, 50
80	2, 4, 6, 8, 10, 20, 30, 50, 80
100	2, 4, 6, 8, 10, 20, 30, 50, 80, 100

- 25 benchmark functions
- 25 runs for each function
- 28125 numbers of experiments
- S parameter $\rightarrow 2$
- The termination criterion \rightarrow the maximum number of function evaluations i.e. $10000 \cdot D$

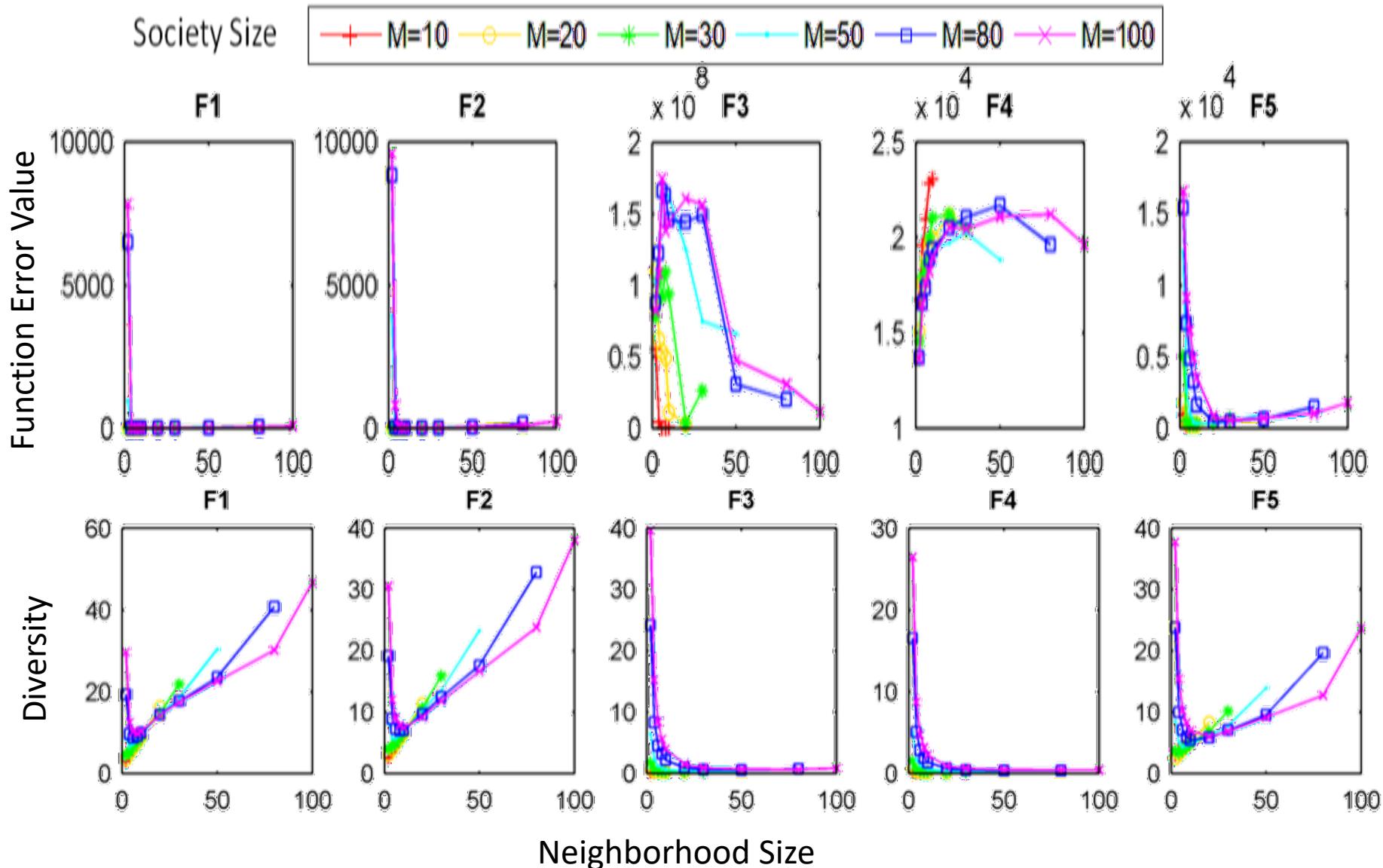
Performance measures

- Function error value of the best individuals for each repetition

$$f(x) - f(x^*)$$

- Diversity of the society, the average of average distance around all the individuals in the society

$$Diversity = \frac{1}{M} \sum_{j=1}^M \left(\frac{1}{M} \sum_{j=1}^M \sqrt{\sum_{k=1}^D (o_{ik} - o_{jk})^2} \right)$$

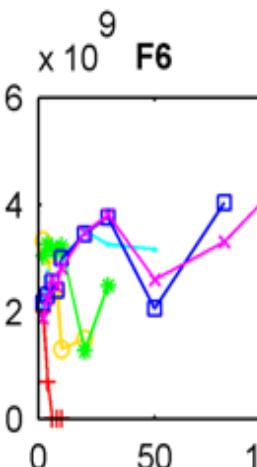


- Unimodal functions do not need highly diverse societies
 - Unimodal functions F1-F5, except for F4 , smaller societies result in lower error values
 - Diversities obtained for smaller societies were lower than the bigger societies
- F4 is noisy in nature and it may disturb the search process.

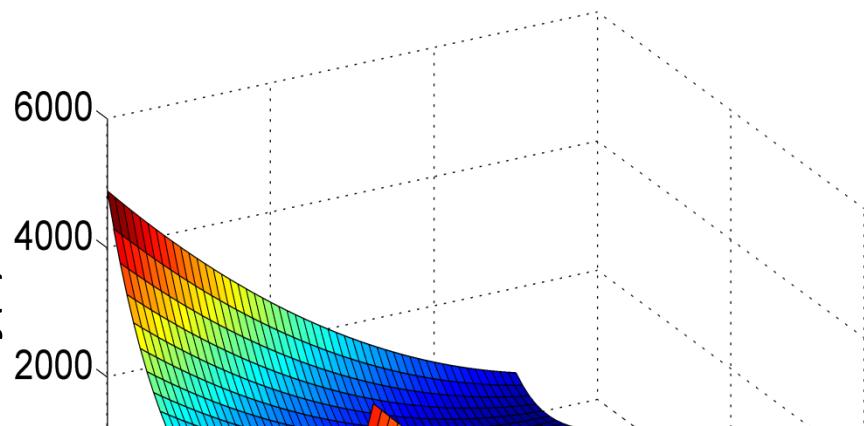
Society Size



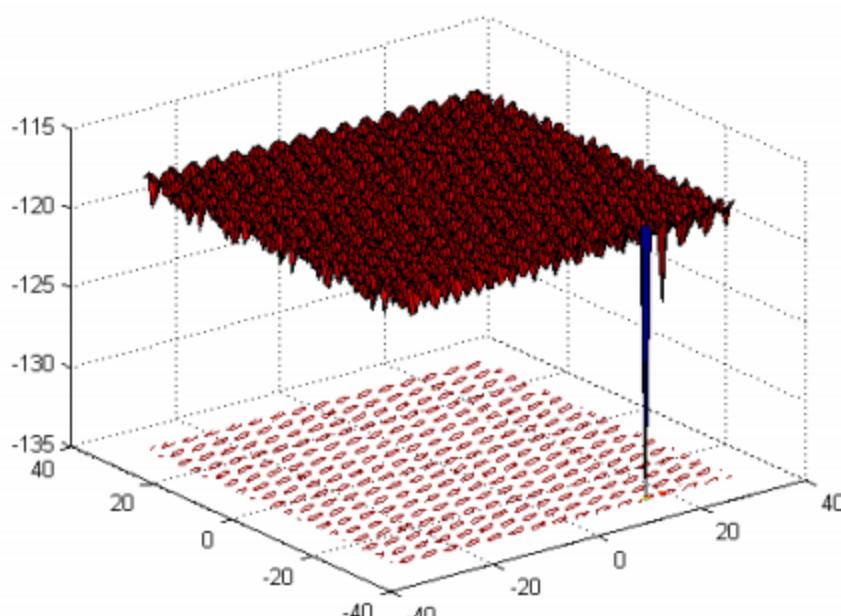
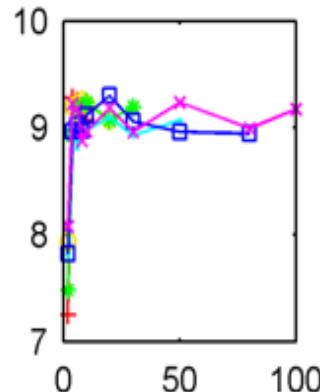
Function Error Value



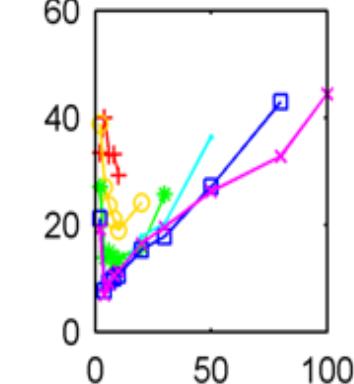
$f(x)$



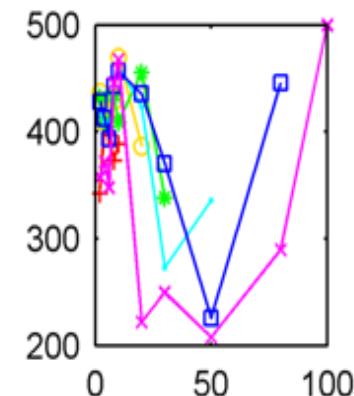
F11



- Multi-modal values.
- F6 , shifted R higher FDC, lower dispersion and gradients
- F8-> needle in a haystack problem and has globally flat structure error values are independent of society size as well as neighborhood size



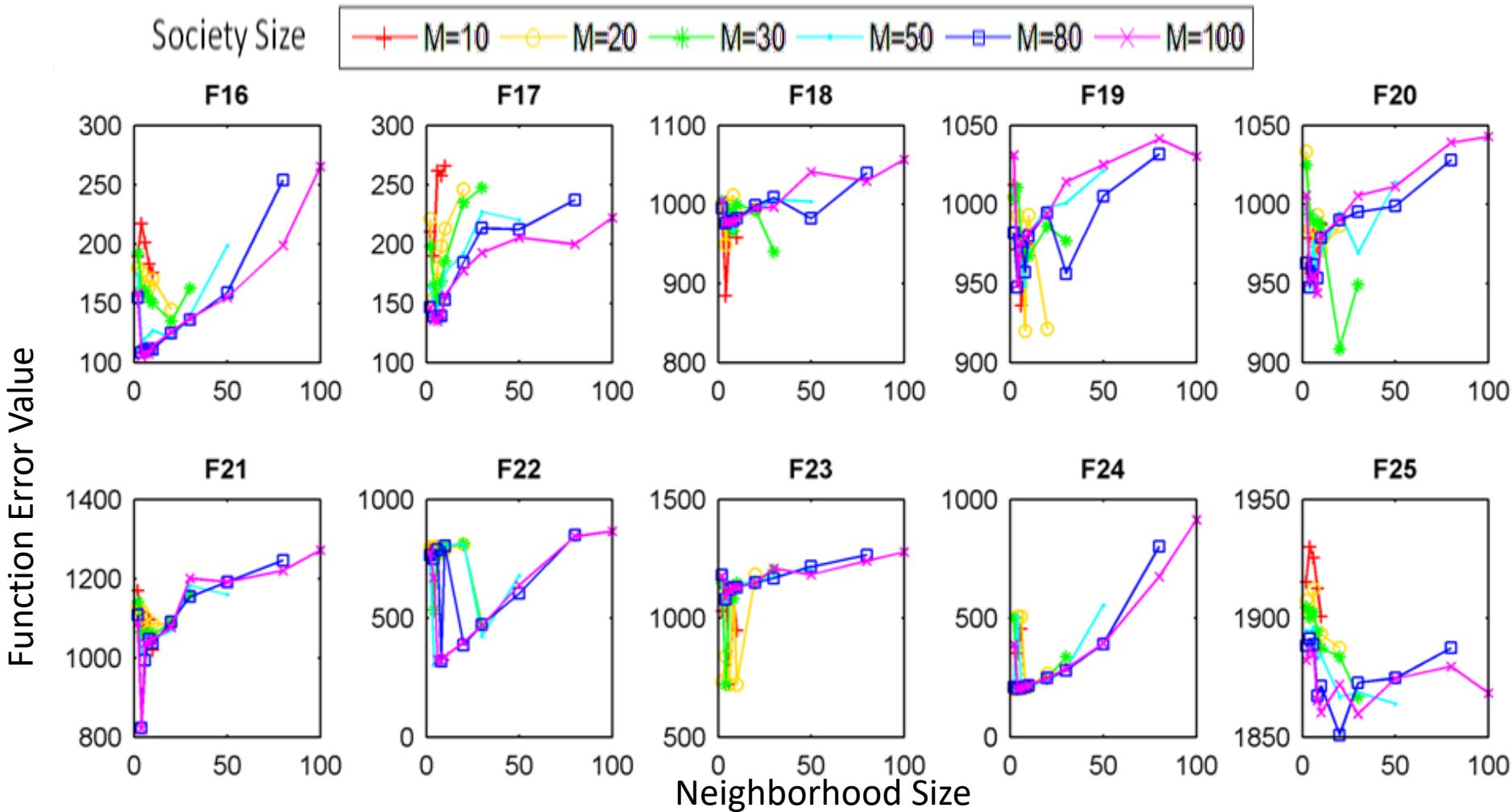
F10



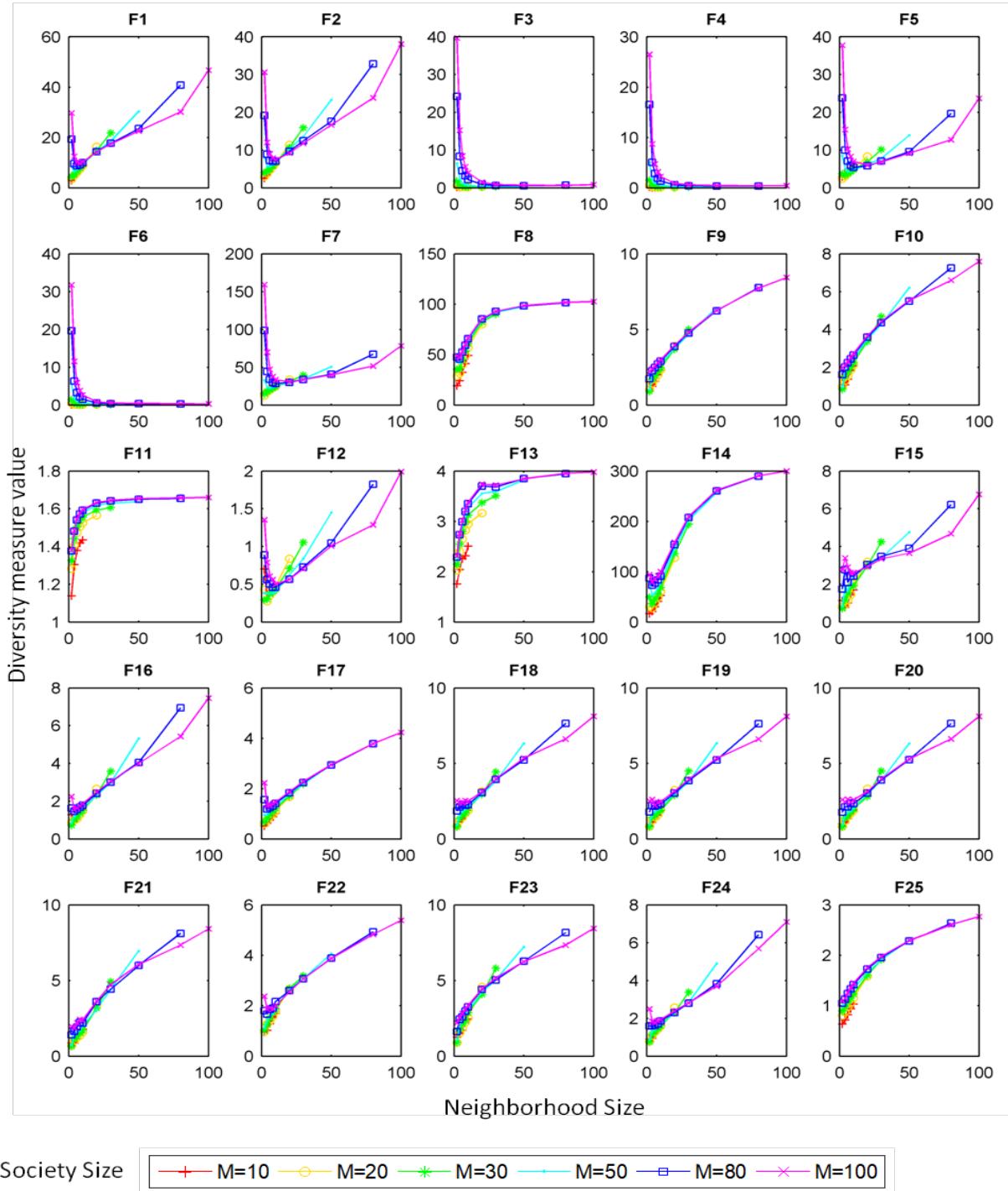
F15

and low error

all length scales,



- Selection of optimum society size and neighborhood size is problem dependent
- The lowest median errors
 - Unimodal functions -> number of neighbors in the range of 4 to 10 and, with small society size (such as $M = 10$)
 - multimodal functions -> large society size (such as $M = 80$ or 100)

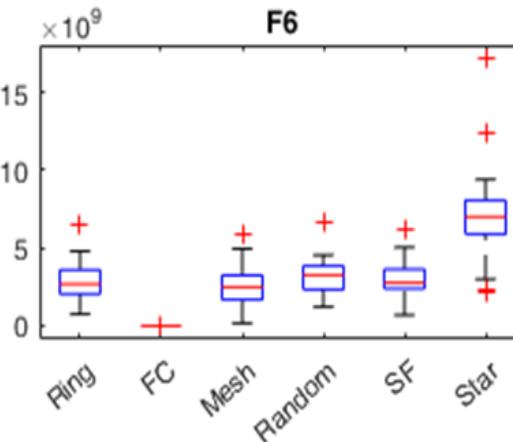
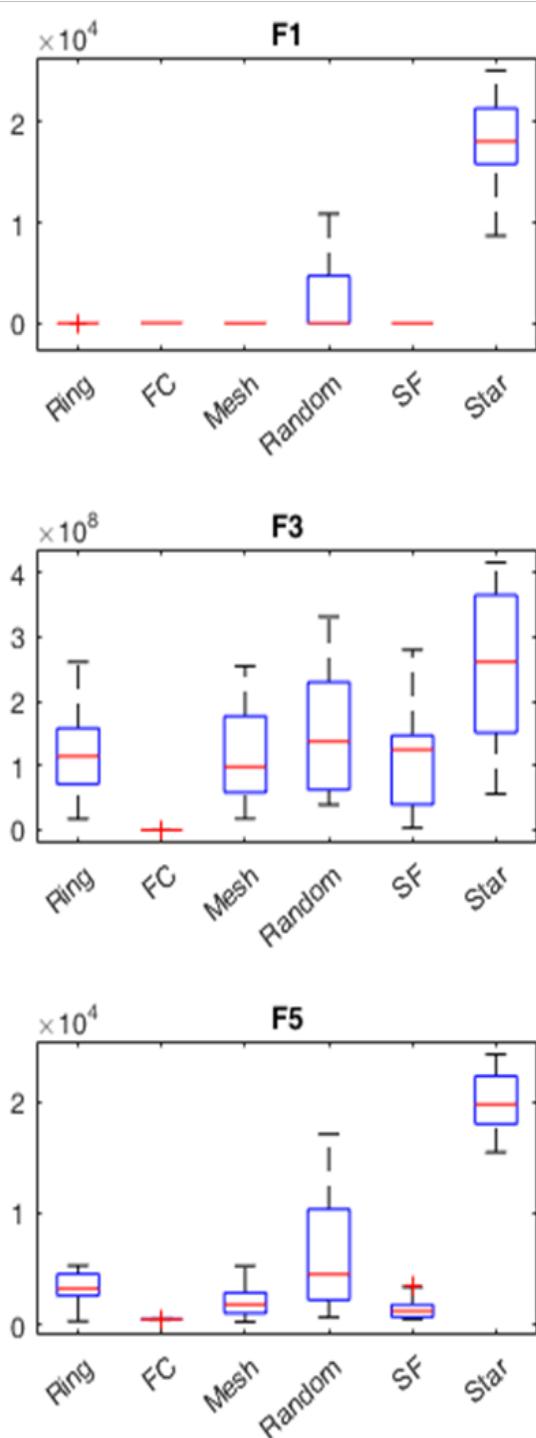
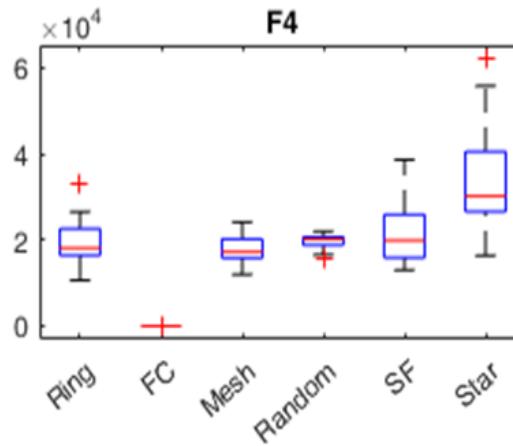
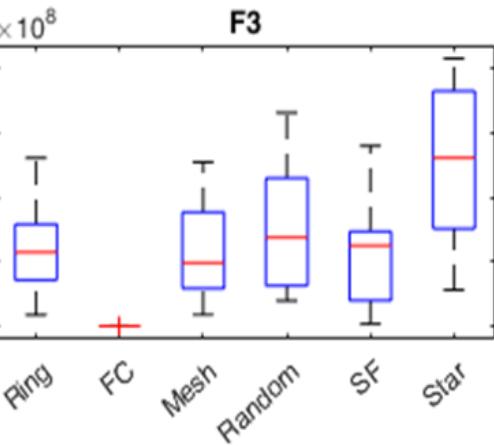
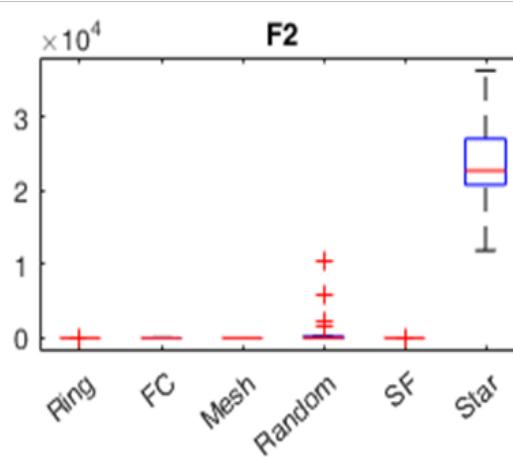
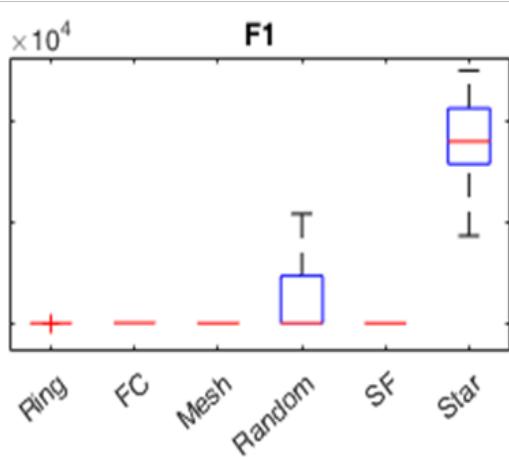


- Different patterns of diversity plots
 - F3, F4, and F6 have L-shaped curves
 - Most of the functions, after reaching a minimum value, diversity curve again starts to rise with the size of neighborhood.
 - F1, F2, F5, F12, and F7 (to some extent) display the boat-shaped behavior, significantly higher diversity values for the extremes of neighborhood size.

Effect of topology type

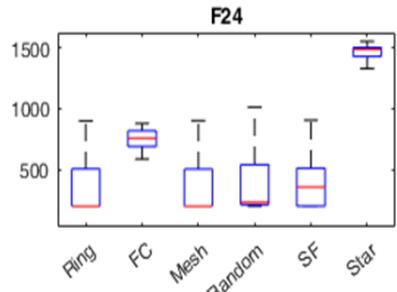
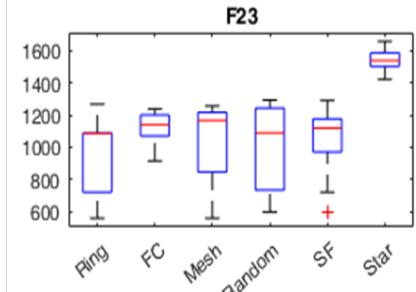
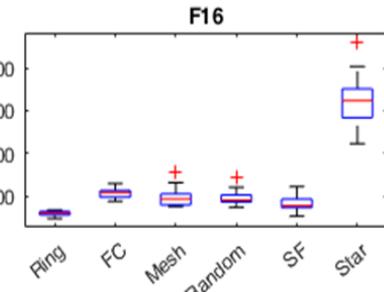
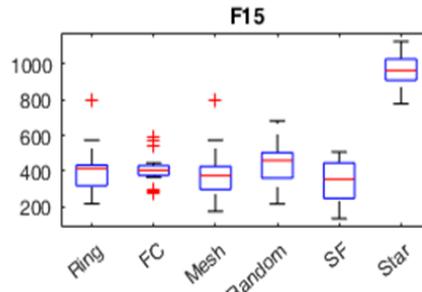
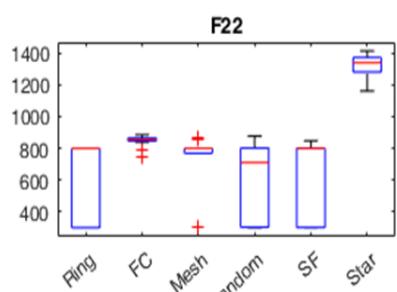
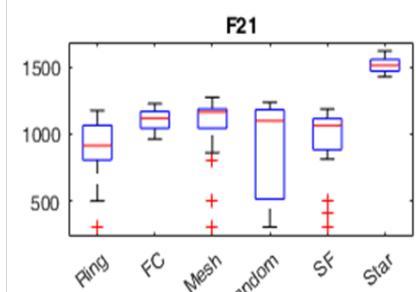
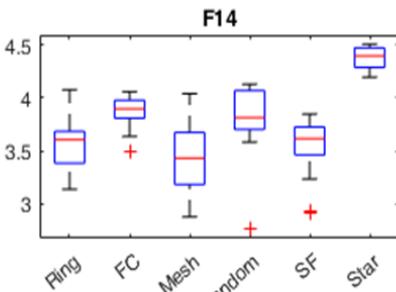
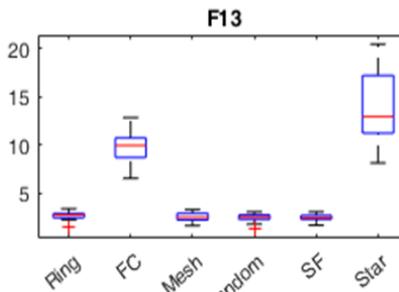
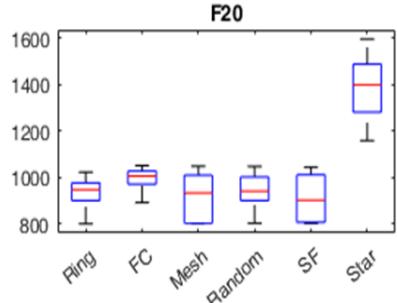
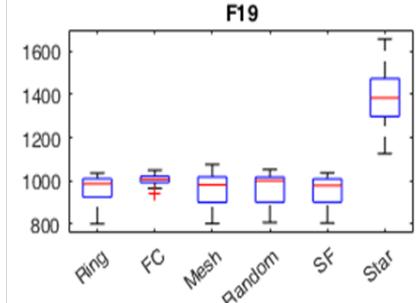
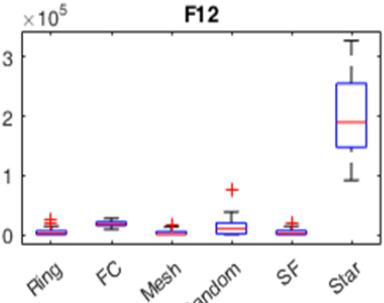
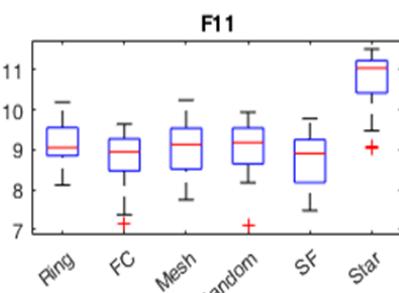
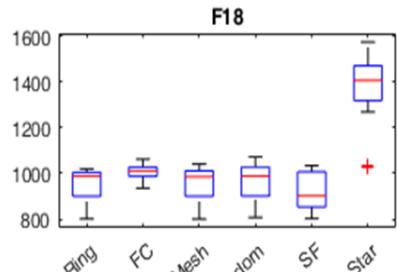
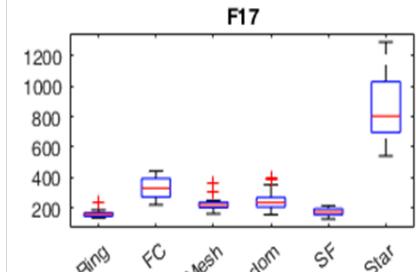
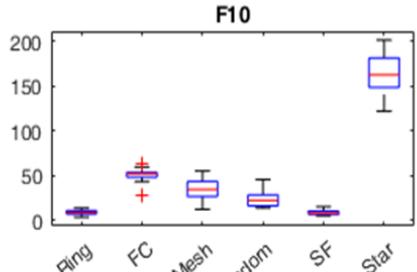
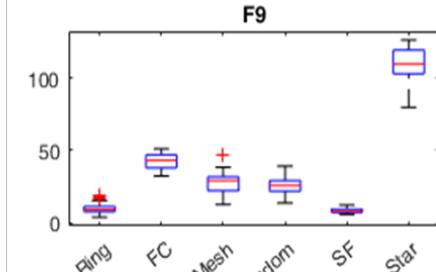
- 6 different topologies
 - **Ring**
 - **Fully connected**
 - **Mesh** : Neumann neighborhood
 - **Random** : Erdos-Renyi model with the p' of 0.0808
 - **Scale-free**: Barabasi and Albert preferential attachment model with m_0 and m set to 2
 - **Star**
- Population size -> 49
- Strength of disintegrating forces -> 2
- Dimensions -> 10
- The average number of neighbors was set to ~4 except star and fully connected topologies

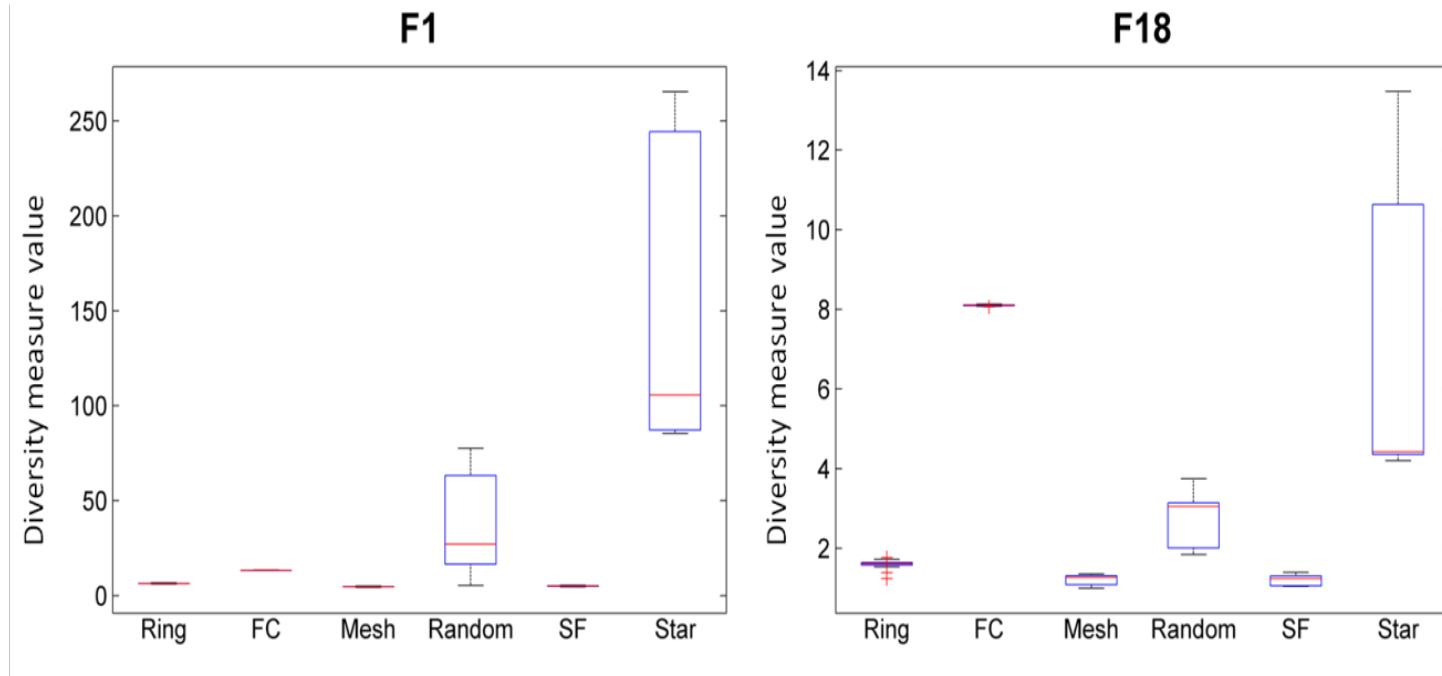
Function error value



- Unimodal function F3-F5 and single tunnel function F6, fully connected topology obtained the minimum error values.
- The information dissemination is the fastest in the fully connected topology which helps in faster convergence to the optimum

Function error value





- The worst performance and the highest diversity values are achieved with star topology.
- It may be because individuals are effectively isolated in the population resulting in higher diversity which affects their ability to work together hence poor performance.

Statistical comparison

- Friedman's non-parametric ranking test with Holm post-hoc procedure

Topology	Ranking	Unadjusted p	p _{Holm}
Ring	2.72	0.289918	0.347235
Fully Connected	3.52	0.010165	0.030496
Mesh	2.88	0.173617	0.347235
Random	3.8	0.001940	0.007759
Scale-free	2.16	---	---
Star	5.92	0	0

- Scale-free is significantly better than the star, random and fully connected whereas; no statistical differences were obtained from ring and mesh

Prediction of pleasantness of the molecules

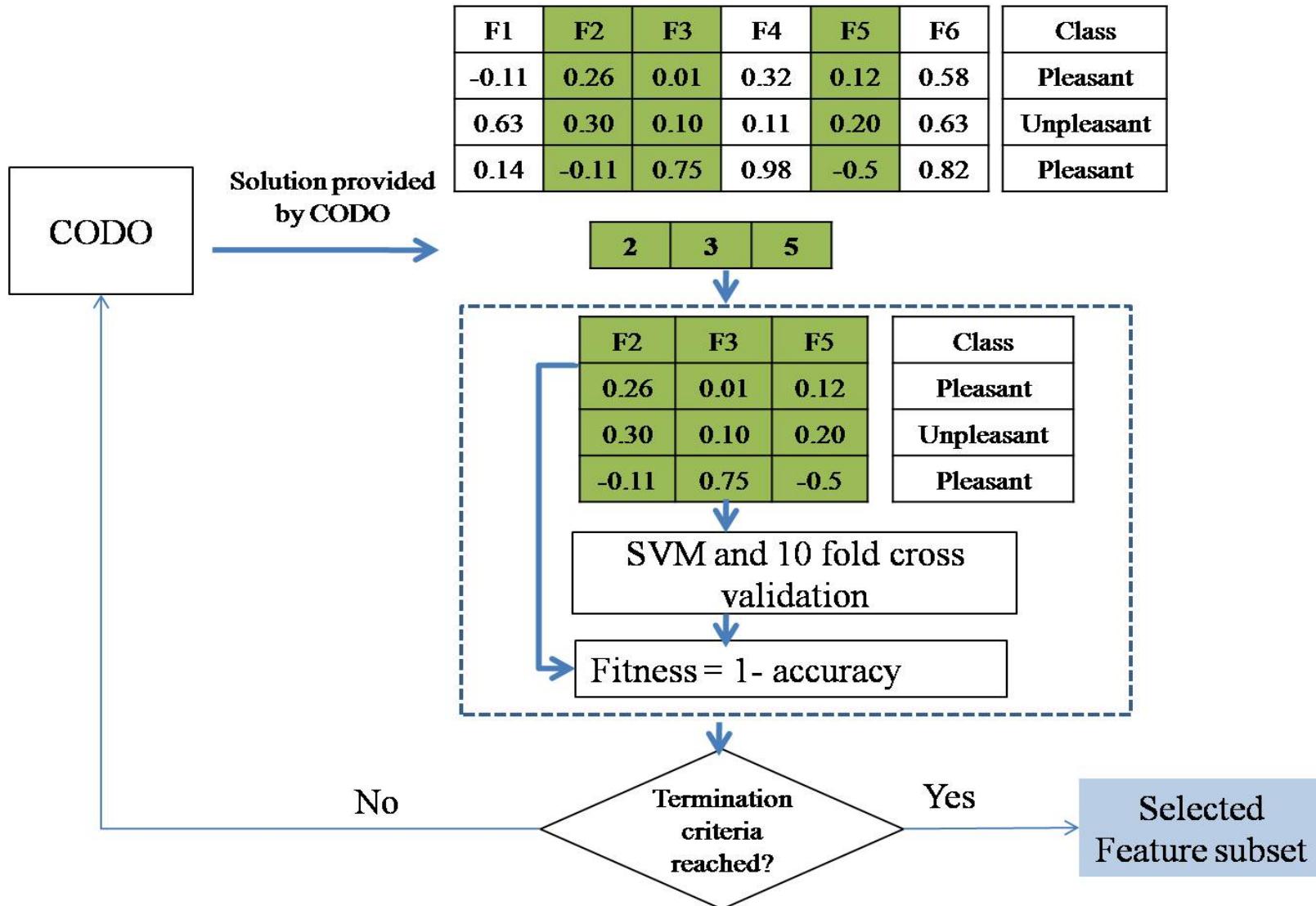
CODO for selecting important feature subset for predicting pleasantness of molecule

Dataset:

- Olfaction challenge by IBM and Rockefeller university
- 49 untrained human subjects were asked to rate 476 odorants
- The subjects rated the compounds between score 0-100, with 0, meaning no observable smell and 100 being the highest
- The molecular descriptor of the compounds was obtained using Dragon (a commercial software package to generate descriptors). *The dataset has 4870 features to relate to the perceptual quality of a molecule.*

Methodology

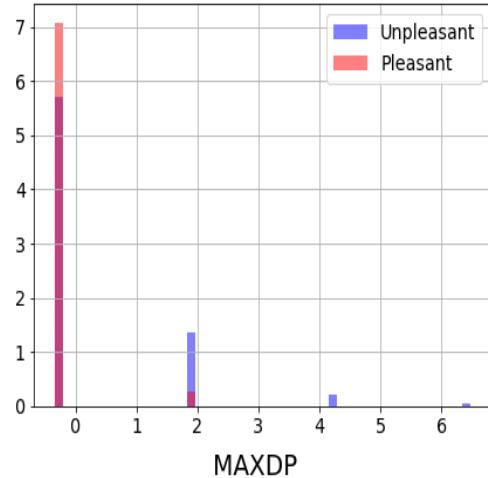
- Removed the features whose standard deviations were zero
 - 3031 physico-chemical features.
 - 1556 miRNA features
- Mean and std deviation normalization
- Split dataset in training (80%) and testing (20%)
- Performed a wrapper based feature selection employing continuous opinion dynamics optimizer and support vector machines.
- Comparison with standard feature selection approaches
 - F-score based feature selection,
 - Support vector machines (SVM) based feature selection
 - Feature transformation using Principal Component Analysis (PCA).



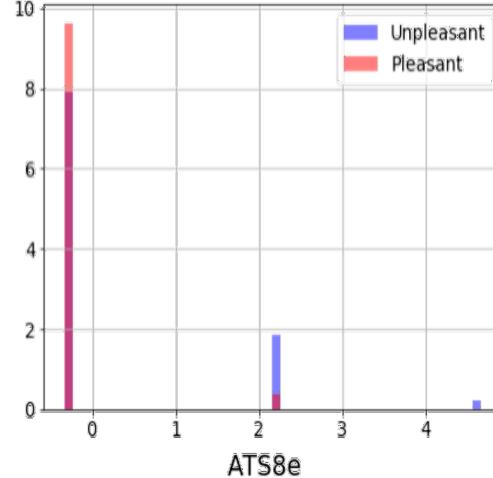
Wrapper approach combining CODO and SVM for feature subset selection

Results: Prediction of pleasantness of a molecule

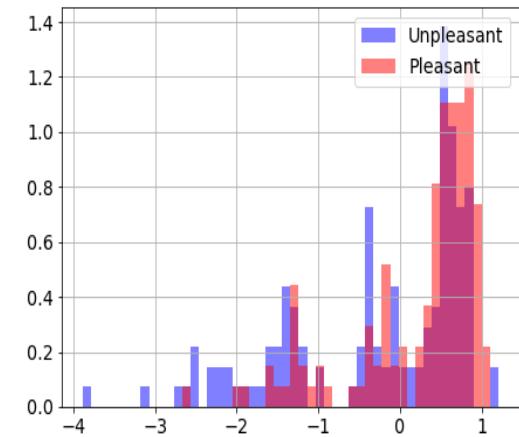
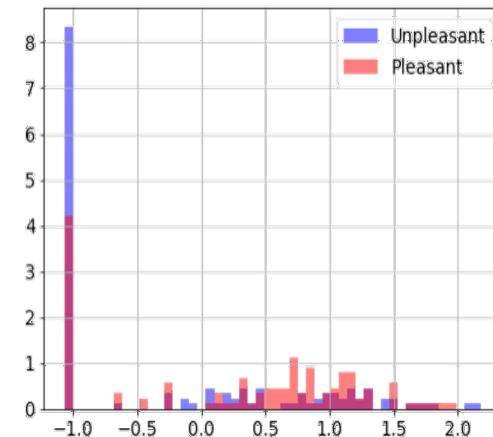
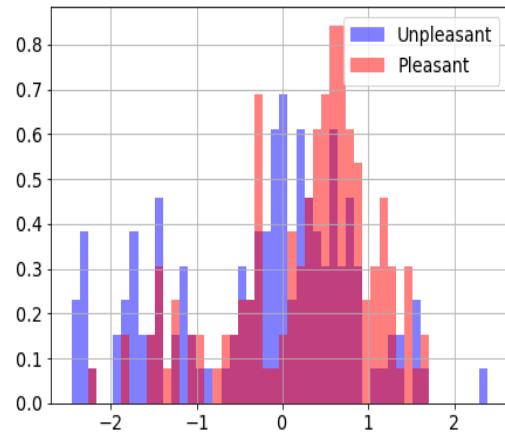
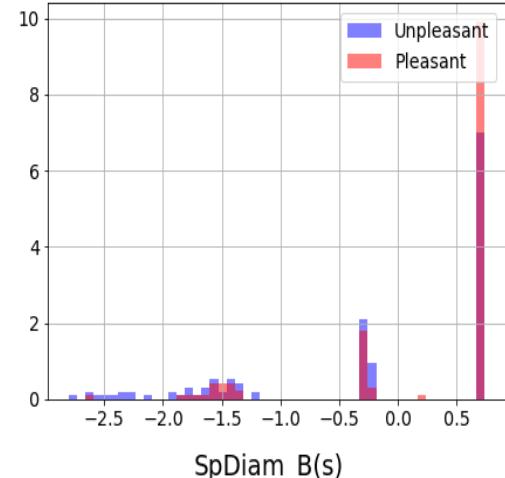
nHM



nS

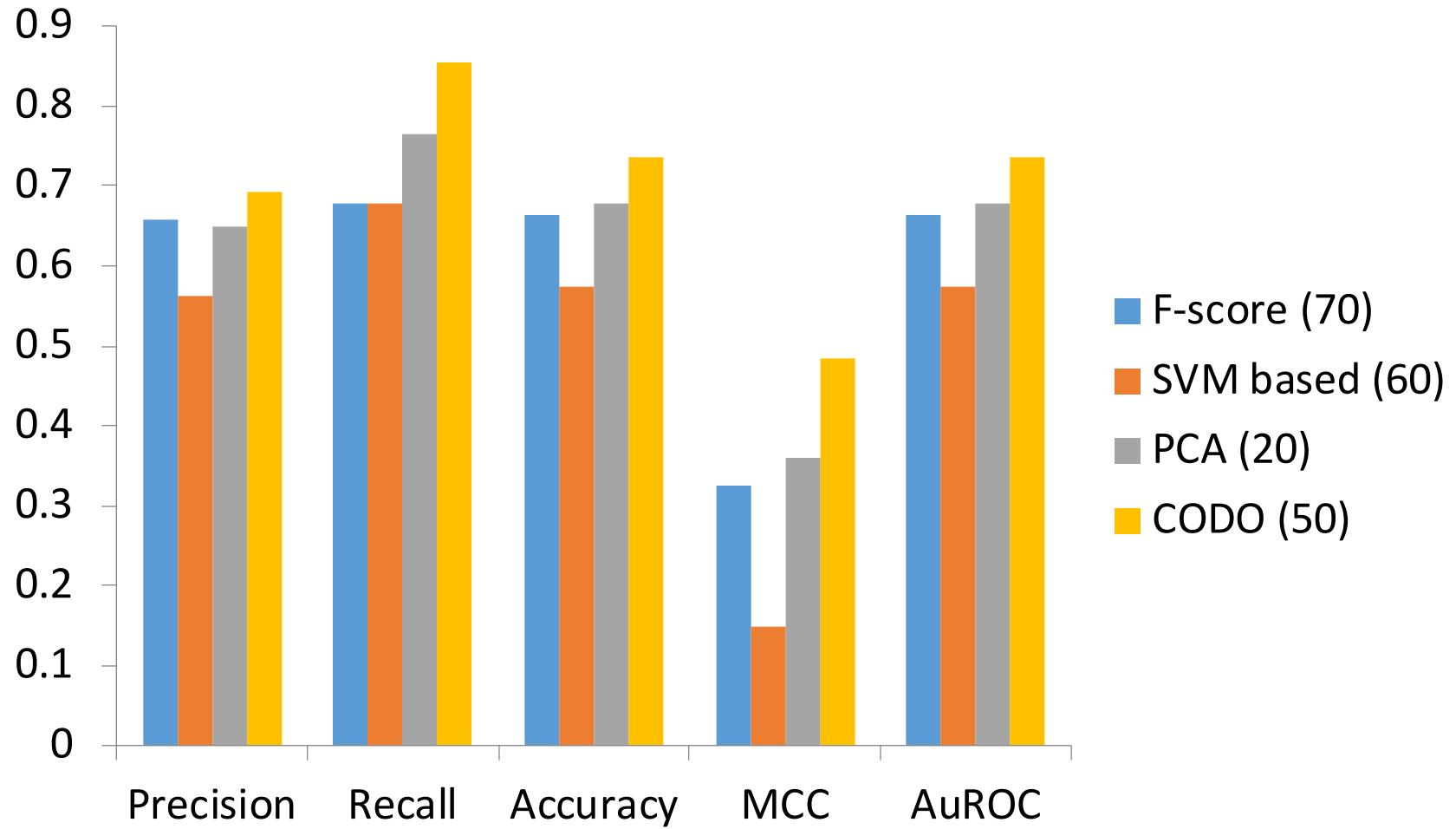


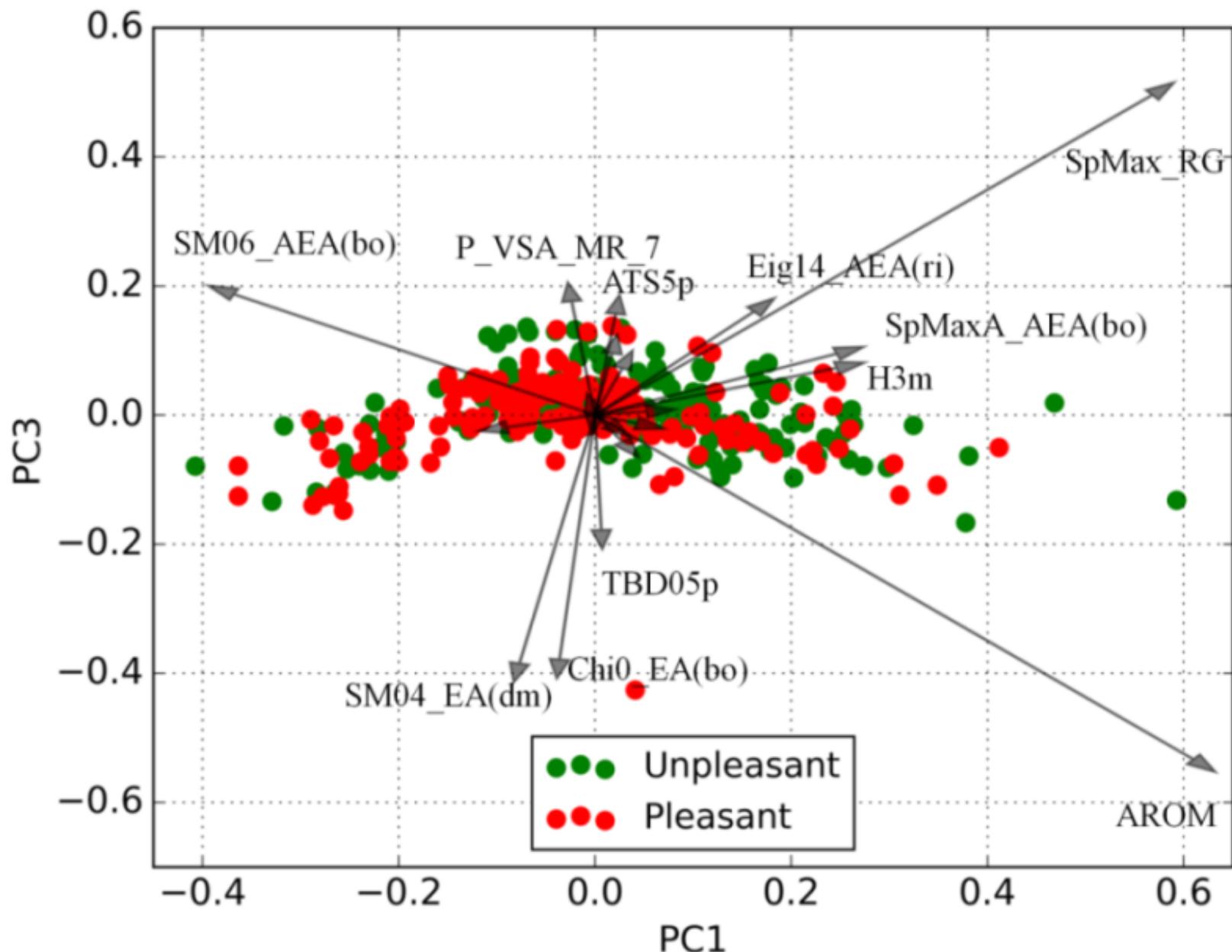
SpMax_B(s)



Top features selected using f-score based method

Results: Prediction of pleasantness of a molecule



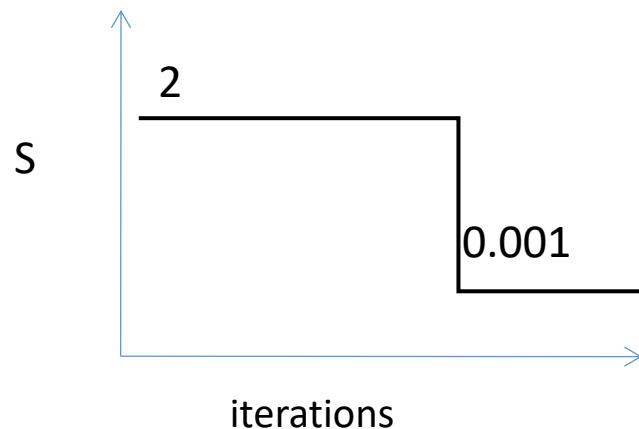


PCA biplot representation (PC1- PC3) of the 20 features selected using wrapper approach based on CODO and SVM for pleasantness dataset

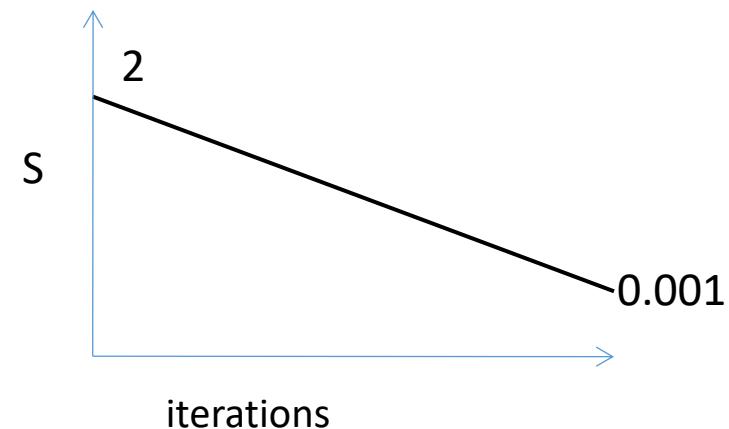
Thank You

Variants of CODO and performance comparisons

- Based on dynamically adjusting S parameter
 - CODO with individuals having different S parameters (D-CODO)
 - CODO with varying S parameter (V-CODO):
 - CODO with linearly varying S parameter (L-CODO)
 - CODO with Quasi-Newton method (QN-CODO)
- Local best particle swarm optimization (PSO)

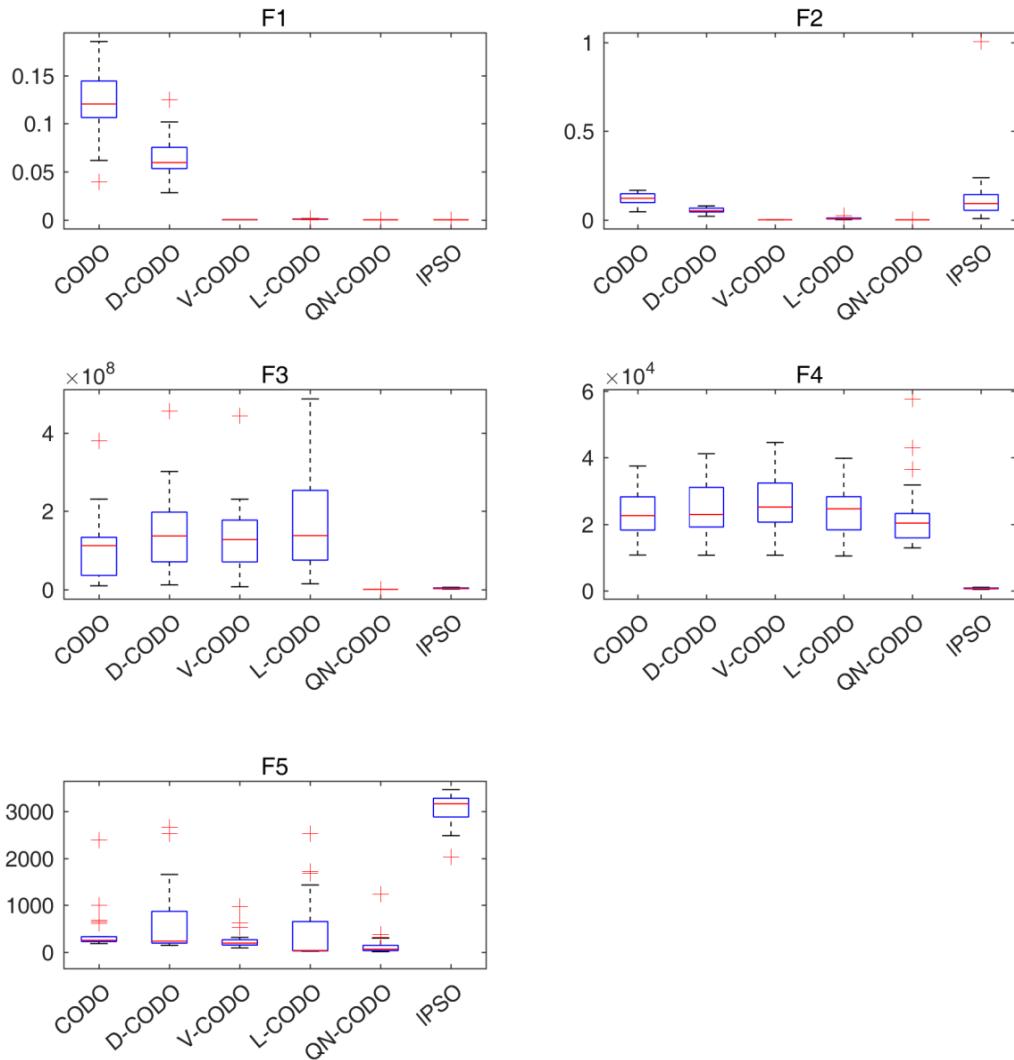


V-CODO



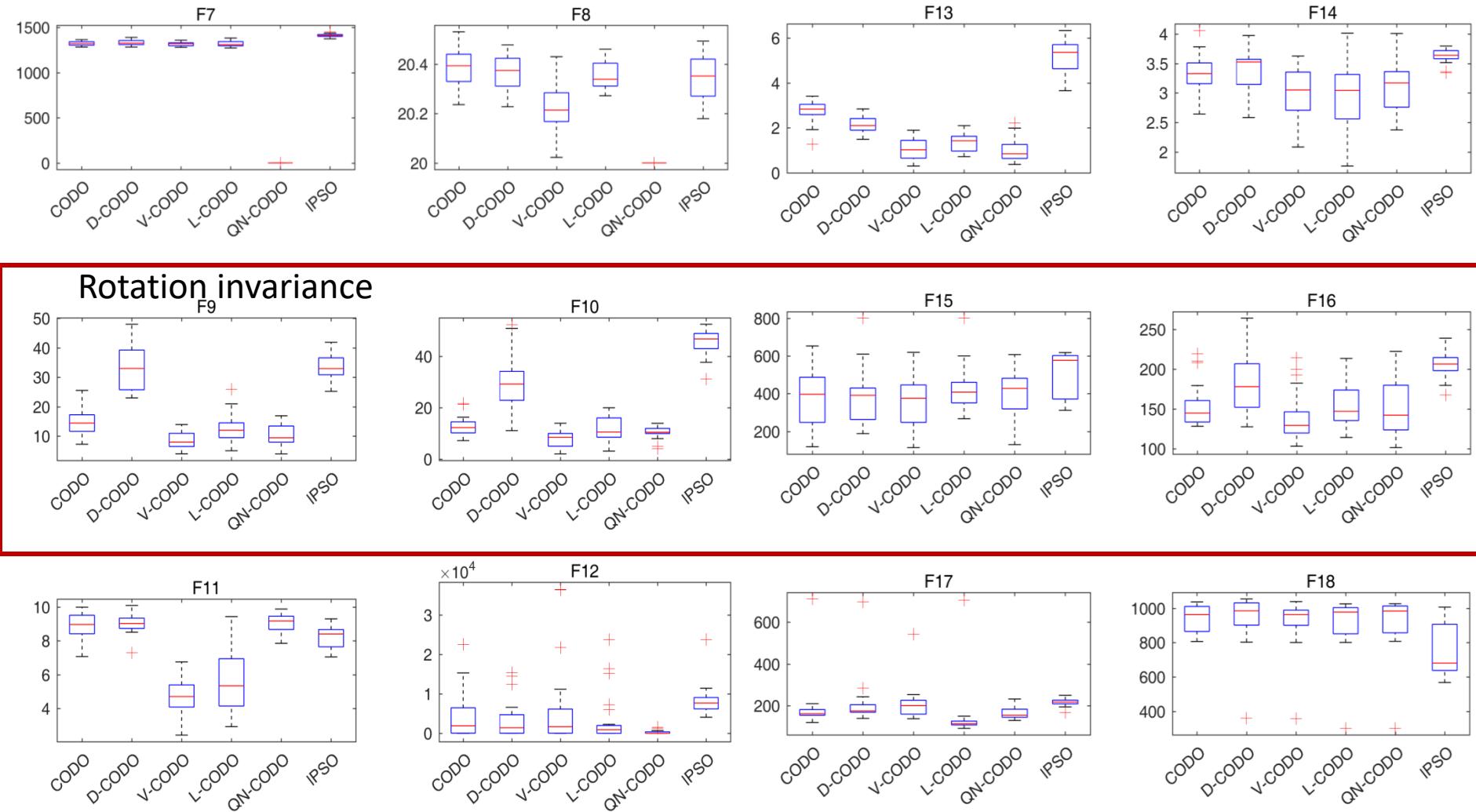
L-CODO

Results: Unimodal functions



- QN-CODO performed the best
 - Also for F3 (ill-conditioned function)
- Effect of addition of noise (F2 & F4):
 - Performance deteriorated

Multi-modal functions



Statistical comparison

Friedman's non-parametric ranking test with Holm post-hoc procedure was performed

Algorithm	Ranking	Unadjusted p	p_{Holm}
CODO	4.04	5.80E-03	2.32E-02
D-CODO	4.58	1.57E-04	7.85E-04
V-CODO	2.92	5.21E-01	9.45E-01
L-CODO	2.96	4.73E-01	9.45E-01
QN-CODO	2.58	---	---
IPSO	3.92	1.13E-02	3.40E-02

- QN-CODO: best ranked algorithm
 - Single funnel functions (F1-F10) with exceptional performance in case of F7 and F8
- D-CODO was the worst performing algorithm

F-Score based feature selection

- Ranked the features according to the f-score or the prediction power of the features.

F-Score = Inter-group variance/Intra-group variance

- After ranking the features according to the f-score, the first 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 features were extracted .
- Further, subjected these features to classification using SVM with linear

$$\text{Inter group variance} = \frac{\sum_{i=1}^K n_i (\bar{X}_i - \bar{X})^2}{K - 1}$$

$$\text{Intra group variability} = \sum_{i=1}^K \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 / (N - K)$$

Support vector machine based feature selection

- The relative importance of each feature is measured by its classification accuracy.
- After calculating the classification performance of each feature, the top r ranked features with large scores are selected.
- Further, subjected these features to classification using SVM with linear

Principal component analysis based feature transformation

- The data is transformed using PCA
- The top r principal components (PCs) in the order of variance are subjected to a classifier to evaluate their performance

Performance measures

		Actual Class	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

- **Recall/ Sensitivity** = $TP / (TP + FN)$
- **Precision** = $TP / (TP + FP)$
- **Accuracy** = $TP + TN / (TP + TN + FP + FN)$
- **Matthews Correlation Coefficient (MCC)**

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- **AuROC:** The area under Receiver Operating Characteristics (ROC) curve