INSTITUT NATIONAL DE RECHERHE EN INFORMATIQUE ET AUTOMATIQUE

Comparison of Multiobjective Evolutionary Algorithms.

Progress report by
Ilya G. Loshchilov

Project-Team TAO
Thème apprentissage et optimisation

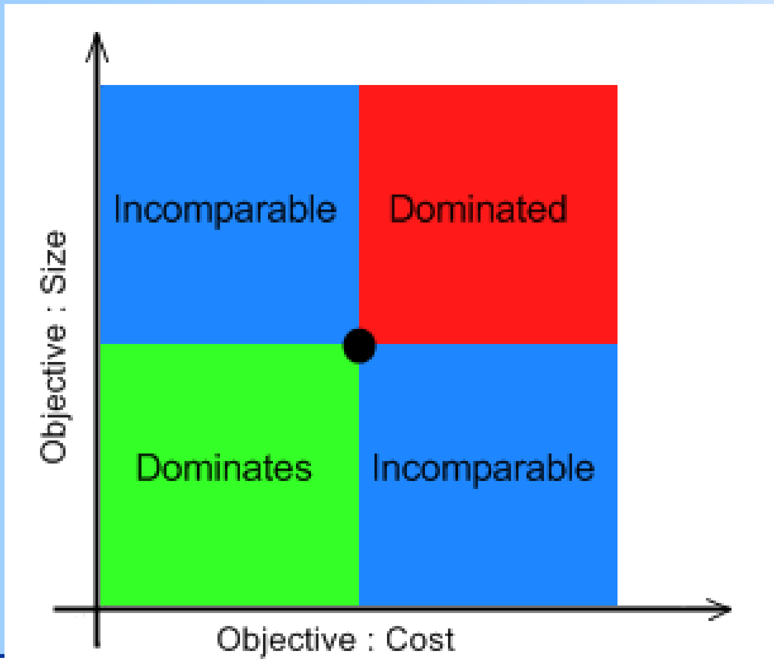
March 9, 2009

Supervisor : Marc Schoenauer

Multiobjective optimization

Many real-word optimization problems involve multiple objectives which are often conflicting. Consequently, instead of a single optimal solution, a set of optimal solutions (called *Pareto-optimal set*) exists for such problems.

The search for an optimal solution has fundamentally changed from what we see in the case of single-objective problems. Each of the Pareto-optimal solutions represents a different compromise between objectives and without preference information, anyone of them no worse than any other.

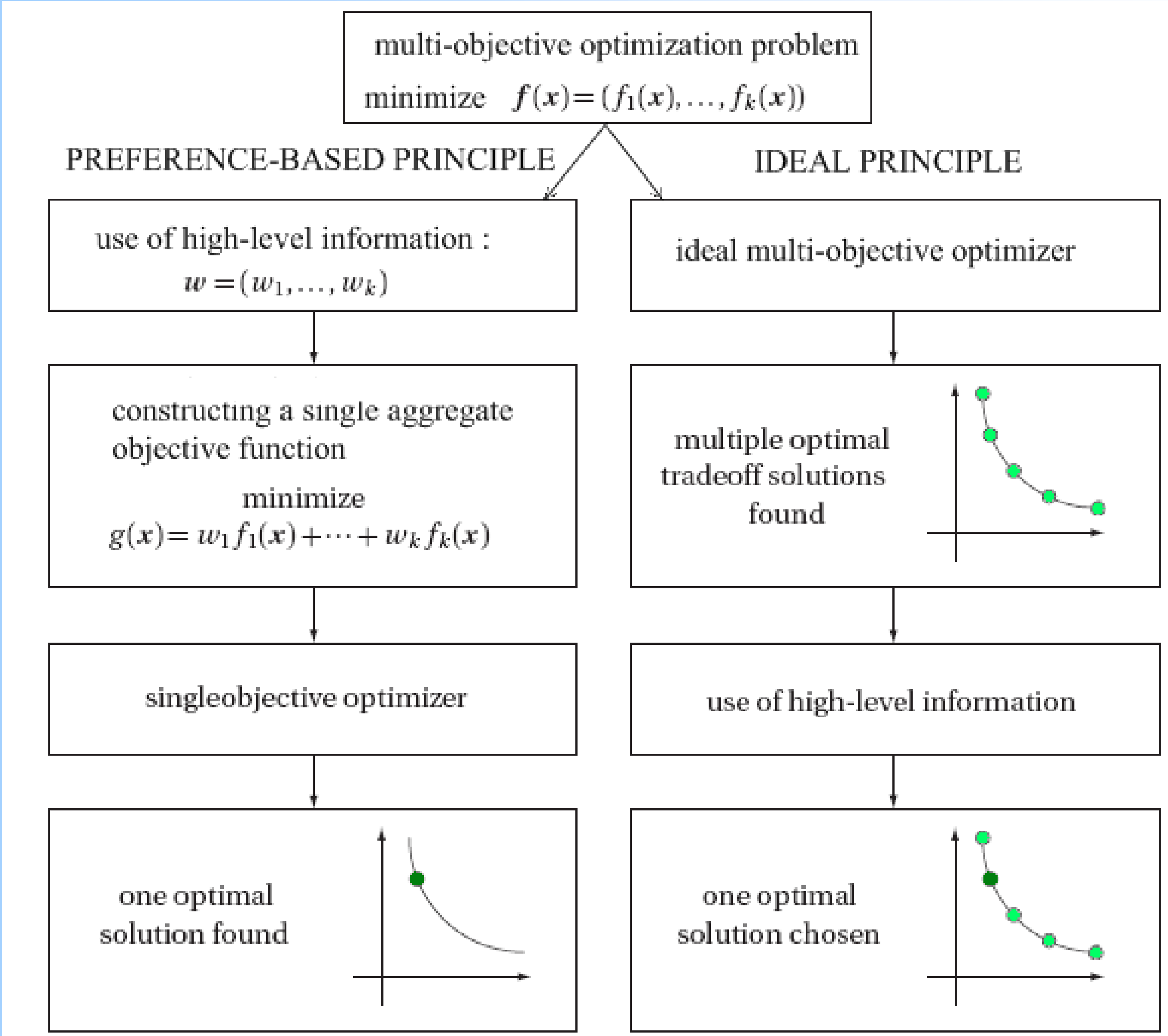


(T.Voss 2007)

Example :
A Cheap and Small Mobile Phone

Chapter 2

Ideal and preference-based principle



(T.Tusar 2007)

Pareto dominance and Pareto optimality

Definition 2.1 (Pareto dominance of vectors). The objective vector z^1 *dominates* the objective vector $z^2\ (z^1 \prec z^2) \overset{def}{\Leftrightarrow} z_j^1 \le z_j^2$ for all $j \in \{1,\dots,m\}$ and $z_k^1 < z_k^2$ for at least one $k \in \{1,\dots,m\}$.

Definition 2.2 (Weak Pareto dominance of vectors). The objective vector z^1 *weakly dominates* the objective vector $z^2\ (z^1 \preceq z^2) \overset{def}{\Leftrightarrow} z_j^1 \le z_j^2$ for all $j \in \{1,\dots,m\}$.

Definition 2.3 (Strict Pareto dominance of vectors). The objective vector z^1 *strictly dominates* the objective vector $z^2\ (z^1 \prec\prec z^2) \overset{def}{\Leftrightarrow} z_j^1 < z_j^2$ for all $j \in \{1,\dots,m\}$.

When $z^1 = f(x^1), z^2 = f(x^2)$ and z^1 (weakly or strictly) dominates z^2, we say that solution x^1 (weakly or strictly) dominates the solution x^2. Note that $z^1 \prec\prec z^2 \Rightarrow z^1 \prec z^2 \Rightarrow z^1 \preceq z^2$.

Definition 2.4 (Pareto Incomparability of vectors). The objective vector z^1 *incomparable* with objective vector $z^2\ (z^1 \| z^2) \overset{def}{\Leftrightarrow} z_j^1 < z_j^2$ for $j \in \{1,\dots,m\}$ and $z_k^1 > z_k^2$ for $k \ne j$ and $k \in \{1,\dots,m\}$.

Definition 2.5 (Pareto Indifference of vectors). The objective vector z^1 *indifferent* with objective vector $z^2\ (z^1 \sim z^2) \overset{def}{\Leftrightarrow} z_j^1 = z_j^2$ for all $j \in \{1,\dots,m\}$.

Definition 2.6 (Pareto optimality). The solution x^* and its corresponding objective vector $z^* = f(x^*)$ are *Pareto optimal* $\overset{def}{\Leftrightarrow}$ there no exists $z \in Z$ such that $z \prec z^*$.

All Pareto-optimal solutions compose the *Pareto-optimal set*, while the corresponding objective vectors constitute the *Pareto-optimal front*.

"A solution to a MOP is Pareto-optimal if there exists no other feasible solution which would decrease some criteria without causing a simultaneous increase in at least one other criterion." (Coello 2006)

2.2 Assessment of multiobjective optimizers

Two important criteria :

i) the quality of obtained solutions and (diversity, optimality)

ii) the computational cost required to produced them (CPU and NFuncEvals)

Platform for assessment :

Programming language independent interface for search algorithms (PISA) available from http://www.tik.ee.ethz.ch/pisa/ (Knowles et al 2006)

2.2.1 Dominance ranking

Source : Optimizer A and B

Question : Who is the best ?

Procedure : 1. $A_1, A_2, \ldots, A_r$ and $B_1, B_2, \ldots, B_r$ where r = number of runs

2. Gather all in collection $C = \{C_1, \ldots, C_{2r}\}$

3. Each approximation set is ranked according to the number of approximation sets that are better than the selected approximation set

4. One of the statistical rank test can then be used to determine if there exists a significant difference between the values of the two sets.

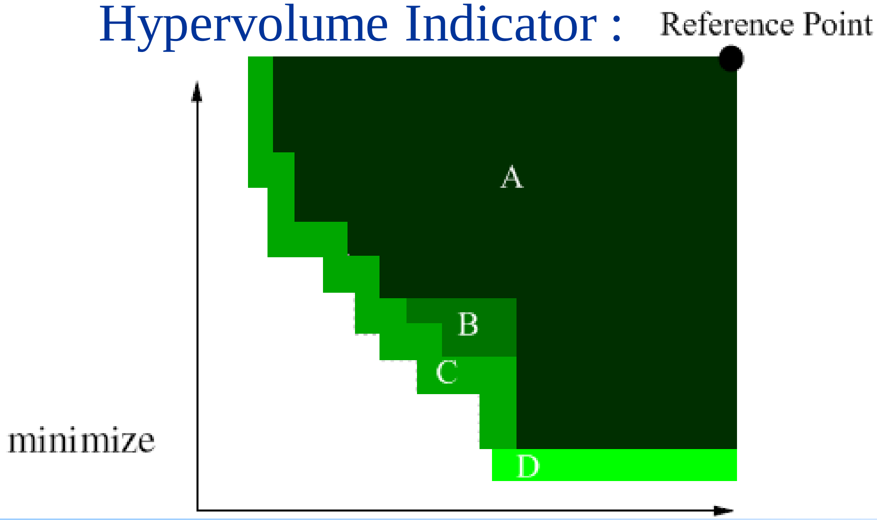
2.2.2 Quality Indicators

Definition 2.12 (Unary quality indicator). The function $I:\Omega \to \mathbb{R}$, which assigns a real value to any approximation set $Z \in \Omega$, is called *unary quality indicator*.

Definition 2.13 (Binary quality indicator). The function $I:\Omega\times\Omega \to \mathbb{R}$, which assigns a real value to any approximation set $(Z_1, Z_2) \in \Omega\times\Omega$, is called *binary quality indicator*.

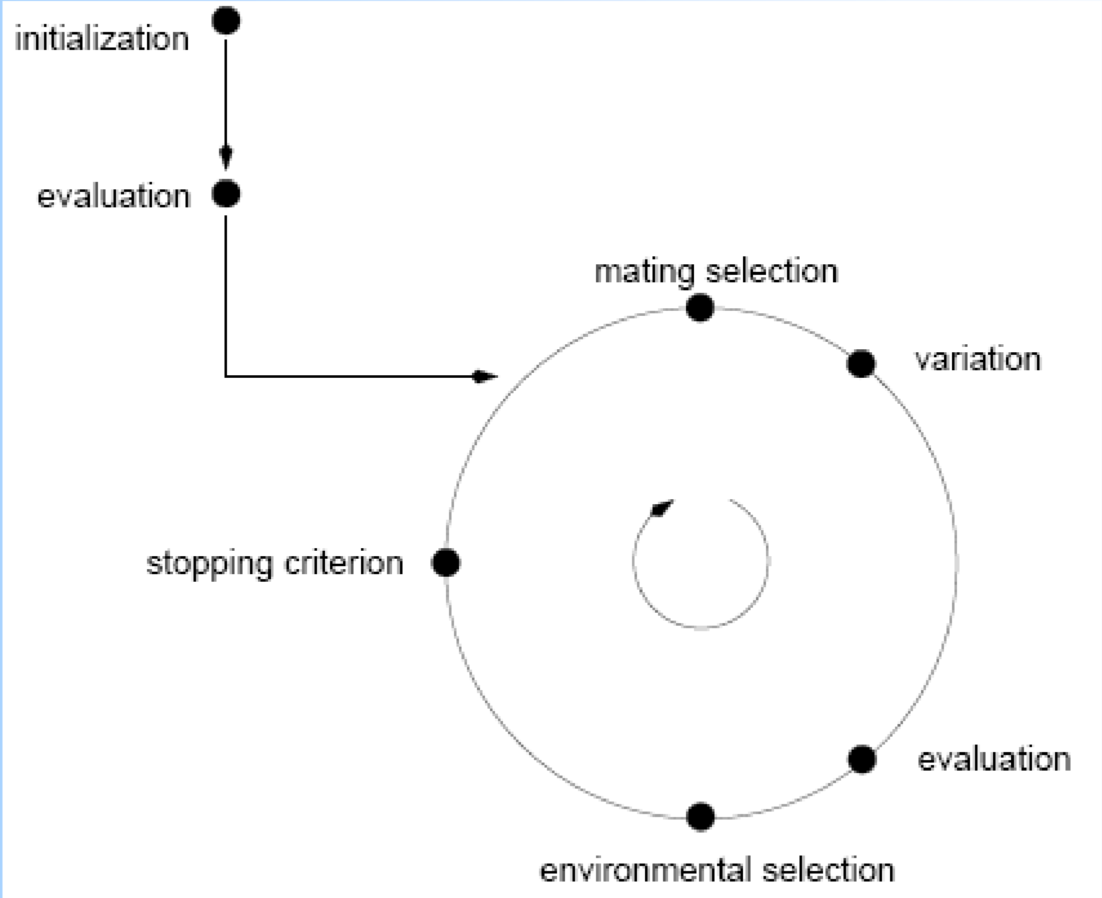
Definition 2.14 (Pareto compliant indicator). The unary indicator $I:\Omega \to \mathbb{R}$ is *Pareto compliant* $\overset{def}{\Leftrightarrow}$ for every pair of approximation sets Z_1 and Z_2 for which $Z_1 \prec Z_2$ and $I(Z_1)$ is not worse than $I(Z_2)$.

Hypervolume Indicator :



The binary additive epsilon indicator I(A,B), gives the minimum summand **eps** to which each vector from **B** can be added in every objective such that resulting approximation set is weakly dominated by **A**.

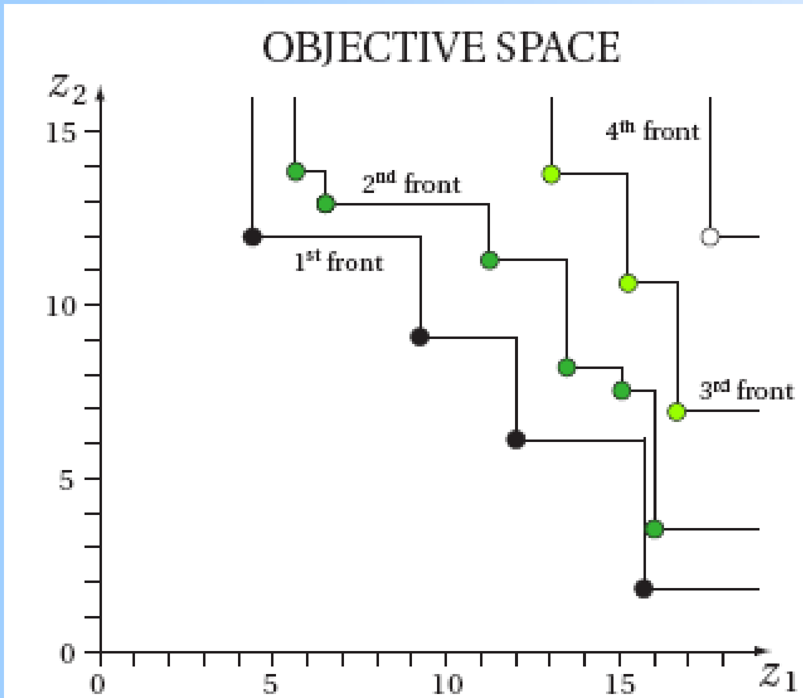
Canonical Multiobjective Evolutionary Algorithms



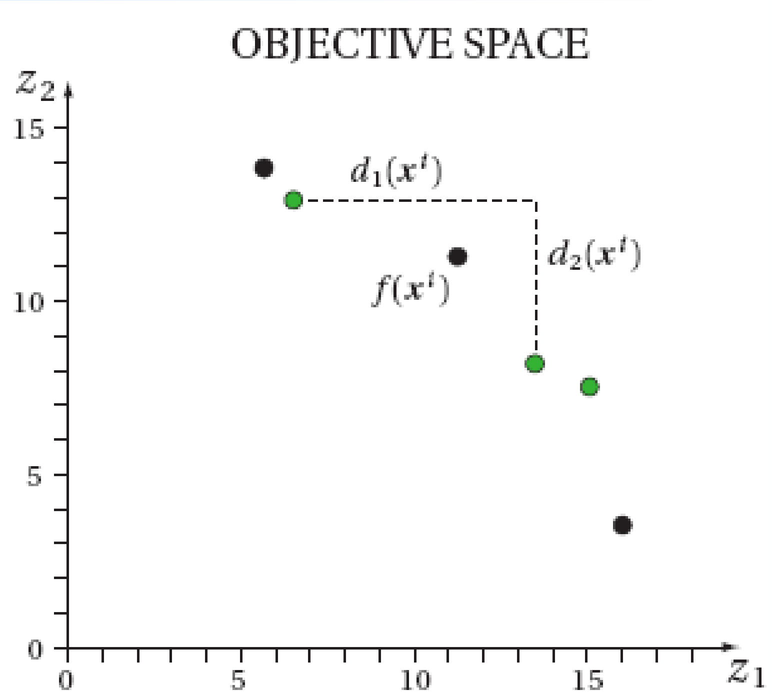
(T.Voss 2007)

NSGA-2, SPEA2, IBEA, Epsilon-MOEA …

Nondominated sorting (Deb at el 2002)



(a) nondominated sorting



(b) crowding distance

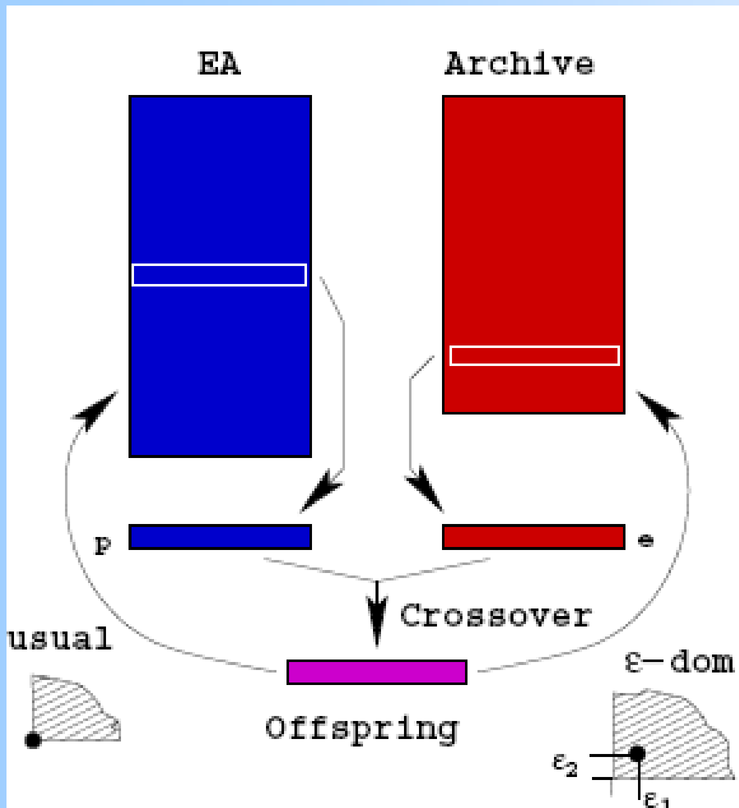
Distance between neighbouring :

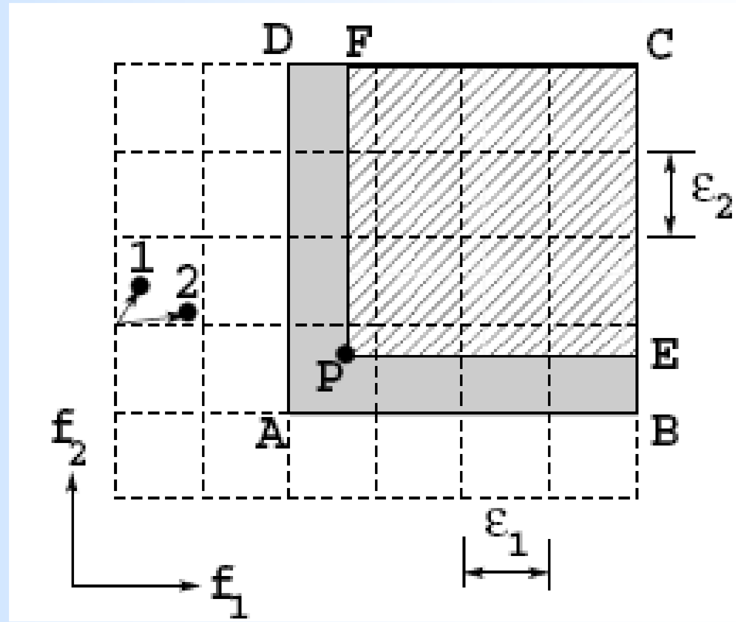
$$d_j(x^i) = \frac{f_j(x^{i+}) - f_j(x^{i-})}{f_j^{\max} - f_j^{\min}}$$

Crowding distance :

$$c(x^i) = \sum_{j=1}^{m} d_j(x^i)$$

The epsilon-Dominance Concept (Deb 2003)



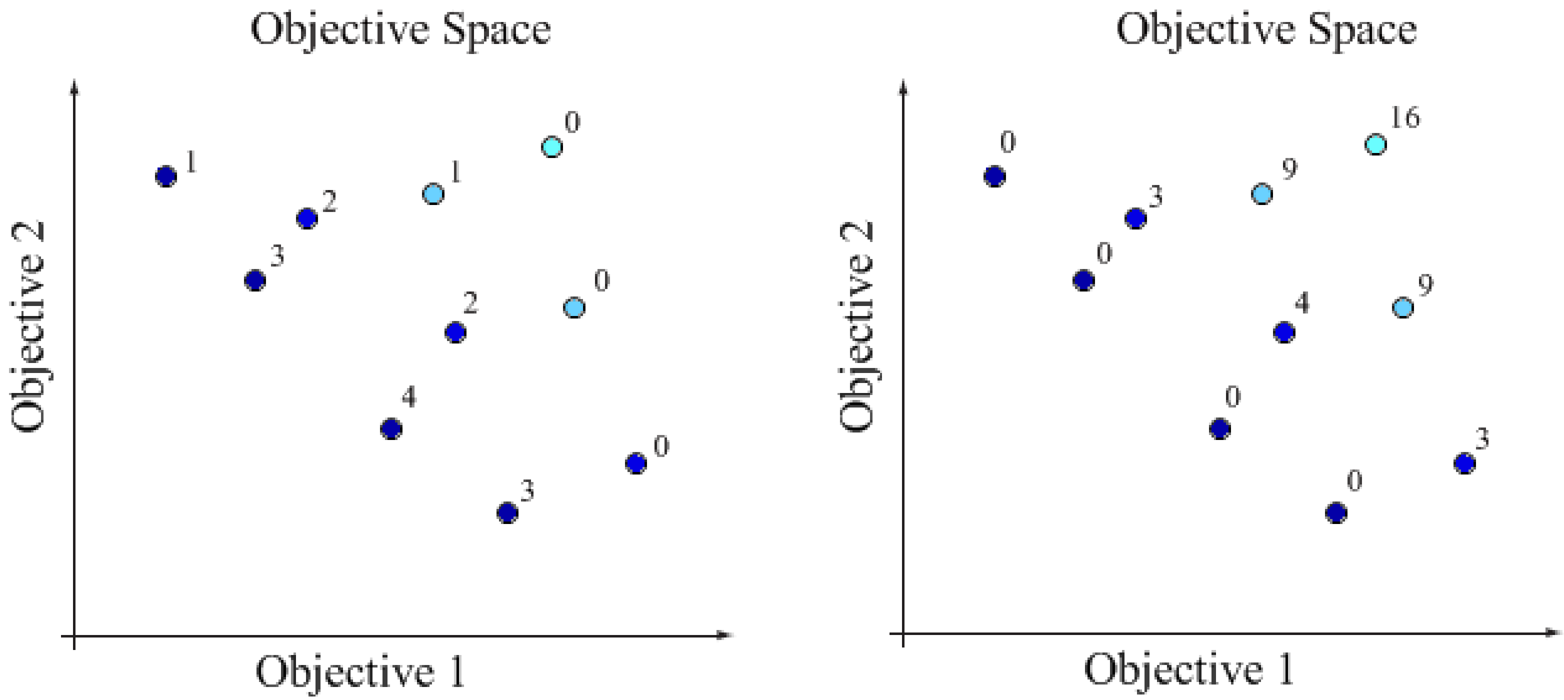


$$B_j(f) = \begin{cases} \left|(f_j - f_j^{\min})/e_j\right|, & \textit{for Minimizing } f_j \\ \left|(f_j^{\max} - f_j)/e_j\right|, & \textit{for Maximizing } f_j. \end{cases}$$

Strength Pareto approach (Zitzler et al 2001)

The strength of an individual at generation t is equal to the number of individuals from A_{t-1} and Q_{t-1} that are dominated by it: $S(x^i) = |\{x^j \in A_{t-1} \cup Q_{t-1} \mid x^i \prec x^j\}|$.

The raw fitness of an individual is computed by summing the strengths of all individuals that dominate it (see Figure 2.3.8 (b)): $R(x^i) = |\{x^j \in A_{t-1} \cup Q_{t-1} \mid x^i \prec x^j\}|$.



After several generations the majority of individuals become nondominated (and have raw fitness equal to 0), therefore addition information must be used to obtain spread in the objective space. For all individuals x^i with the same raw fitness $R(x^i)$, the density is calculated as:

$$D(x^i) = \frac{1}{\sigma_i^k + 2}, \qquad F(x^i) = R(x^i) + D(x^i).$$

Indicator-Based selection (Zitzler et al 2004)

"The main idea is to first define the optimization goal in terms of a binary performance measure (indicator) and then to directly use this measure in selection process."(Zitzler 2004).

Every generation , the objective values of all individuals must be normalized to the [0,1] interval.

Each individual x^1 be evaluated by summing up its indicator values with respect to the rest of population :

$$F'(x^1) = \sum_{x^2 \in R_{t-1} \setminus \{x^1\}} I(\{x^2\},\{x^1\}).$$

The fitness value , which is to be maximized, is a measure for the "loss in quality" if is removed from the population . But IBEA use slightly different scheme , which amplifies the influence of dominating population members over dominated ones

$$F'(x^1) = \sum_{x^2 \in R_{t-1} \setminus \{x^1\}} -e^{I(\{x^2\},\{x^1\})/(ck)},$$

where **k** is a positive scaling factor depending on *I* and *c* is the maximum absolute value of *I* on individuals from R_{t-1}. The fitness defined in this way should be minimized.

Covariance Matrix Adaptation for Multiobjective Optimization. (Igel et al 2006)

Original MO-CMA-ES also names as $\lambda_{MO} \times (1+1)$ -MO-CMA-ES, because of a population of λ_{MO} elitist (1+1)-CMA-ES (Hansen et al 2001). The kth individual in generation g is denoted by

$a_k^{(g)} = [x_k^{(g)}, \bar{p}_{succ,k}^{(g)}, \sigma_k^{(g)}, \bar{p}_{c,k}^{(g)}, C_k^{(g)}]$, where

$x_k^{(g)}$ is the current search point ,

$\bar{p}_{succ,k}^{(g)}$ is the smoothed success probability,

$\sigma_k^{(g)}$ is the global step-size,

$\bar{p}_{c,k}^{(g)}$ is the cumulative evolution path

$C_k^{(g)}$ is the covariance matrix of the search distribution

There are two variants of original MO-CMA-ES : the *c*-MO-CMA-ES and *s*-MO-CMA-ES, which use the crowding-distance and the contributing hypervolume as second level sorting criterion, respectively .

Hypervolume: a is better that *a'* when compared using if either *a* has a better level of non-dominance or *a* and *a'* are on the same level but *a* contributes more to the hypervolume when considering the points at that level of non-dominance.

The contribution hypervolume of an objective vector *a* is the portion of objective space exclusively weakly dominated by *a* .

Chapter 3 Comparison between NSGA2, SPEA2, IBEA, Epsilon-MOEA and MO-CMA-ES

3.1 Benchmark Problems: ZDT (Zietzler et al 2001)

Problem	n	Variable bound	Objective function	Optimal solution
ZDT1	30	[0,1]	$f_1(x) = x_1$ $f_2(x)=g(x)[1-\sqrt{x_1/g(x)}]$, $g(x) = 1+9(\sum_{i=2}^{n} x_i)/(n-1)$	$x_1 \in [0,1]$ $x_i=0$ $i = 2,..n$
ZDT2	30	[0,1]	$f_1(x) = x_1$ $f_2(x)=g(x)[1-(x_1/g(x))^2]$ $g(x) = 1+9(\sum_{i=2}^{n} x_i)/(n-1)$	$x_1 \in [0,1]$ $x_i=0$ $i = 2,..n$
ZDT3	30	[0,1]	$f_1(x) = x_1$ $f_2(x)=g(x)[1-\sqrt{x_1/g(x)}-\frac{x_1}{g(x)}\sin(10\pi x_1)]$ $g(x) = 1+9(\sum_{i=2}^{n} x_i)/(n-1)$	$x_1 \in [0,1]$ $x_i=0$ $i = 2,..n$
ZDT4	10	$x_1 \in [0,1]$ $x_i \in [-5,5]$ $i=2,...n$	$f_1(x) = x_1$ $f_2(x)=g(x)[1-\sqrt{x_1/g(x)}]$ $g(x) = 1+10(n-1)+\sum_{i=2}^{n}[x_i^2-10\cos(4\pi x_i)]$	$x_1 \in [0,1]$ $x_i=0$ $i = 2,..n$
ZDT6	10	[0,1]	$f_1(x) = 1-\exp(-4x_1)\sin^6(6\pi x_1)$ $f_2(x)=g(x)[1-(f_1(x)/g(x))^2]$ $g(x) = 1+9[(\sum_{i=2}^{n} x_i)/(n-1)]^{0.25}$	$x_1 \in [0,1]$ $x_i=0$ $i = 2,..n$

IHR benchmark problems to be minimized, $y = Ox$, where $O \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, and $y_{max} = 1/\max_j(|o_{1j}|)$. For the definition of h, h_f and h_g

Problem	n	Variable bound	Objective function	Optimal solution
IHR1	30	[-1,1]	$f_1(x) = \lvert y_1 \rvert$	$y_1 \in [0, y_{max}]$
			$f_2(x) = g(y) h_f(1 - \sqrt{h(y_1)/g(y)})$	$y_i = 0$
			$g(y) = 1 + 9(\sum_{i=2}^{n} h_g(y_i))/(n-1)$	$i = 2, \ldots n$
IHR 2	30	[-1,1]	$f_1(x) = \lvert y_1 \rvert$	$y_1 \in [-y_{max}, y_{max}]$
			$f_2(x) = g(y) h_f(1 - (y_1/g(y))^2)$	$y_i = 0$
			$g(y) = 1 + 9(\sum_{i=2}^{n} h_g(y_i))/(n-1)$	$i = 2, \ldots n$
IHR 3	30	[-1,1]	$f_1(x) = \lvert y_1 \rvert$	$y_1 \in [0, y_{max}]$
			$f_2(x) = g(y) h_f(1 - \sqrt{h(y_1)/g(y)} - \frac{h(y_1)}{g(y)} \sin(10\pi y_1))$	$y_i = 0$
			$g(y) = 1 + 9(\sum_{i=2}^{n} h_g(y_i))/(n-1)$	$i = 2, \ldots n$
IHR 4	10	[-5,5]	$f_1(x) = \lvert y_1 \rvert$	$y_1 \in [0, y_{max}]$
			$f_2(x) = g(y) h_f(1 - \sqrt{h(y_1)/g(y)})$	$y_i = 0$
			$g(y) = 1 + 10(n-1) + \sum_{i=2}^{n}[y_i^2 - 10\cos(4\pi y_i)]$	$i = 2, \ldots n$
IHR 6	10	[-1,1]	$f_1(x) = 1 - \exp(-4\lvert y_1 \rvert)\sin^6(6\pi y_1)$	$y_1 \in [-y_{max}, y_{max}]$
			$f_2(x) = g(y) h_f(1 - (f_1(x)/g(y))^2)$	$y_i = 0$
			$g(y) = 1 + 9[(\sum_{i=2}^{n} h_g(y_i))/(n-1)]^{0.25}$	$i = 2, \ldots n$

Auxiliary functions

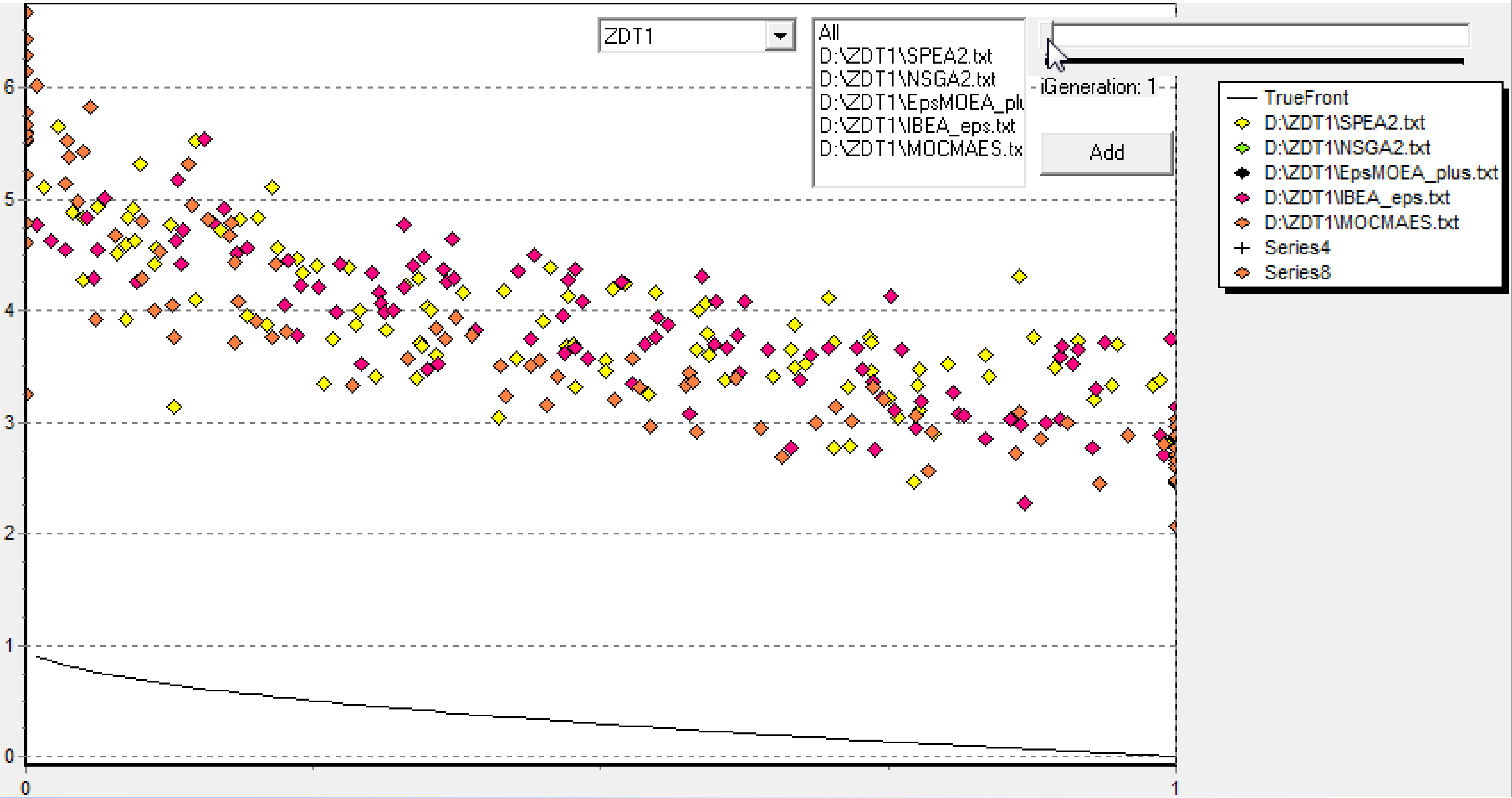
$$h: \quad \mathbb{R} \to [0,1], \quad x \mapsto \left(1 + \exp\left(\frac{-x}{\sqrt{n}}\right)\right)^{-1}$$

$$h_f: \quad \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} x & \text{if } \lvert y_1 \rvert \le y_{max} \\ 1 + \lvert y_1 \rvert & \text{otherwise} \end{cases}$$

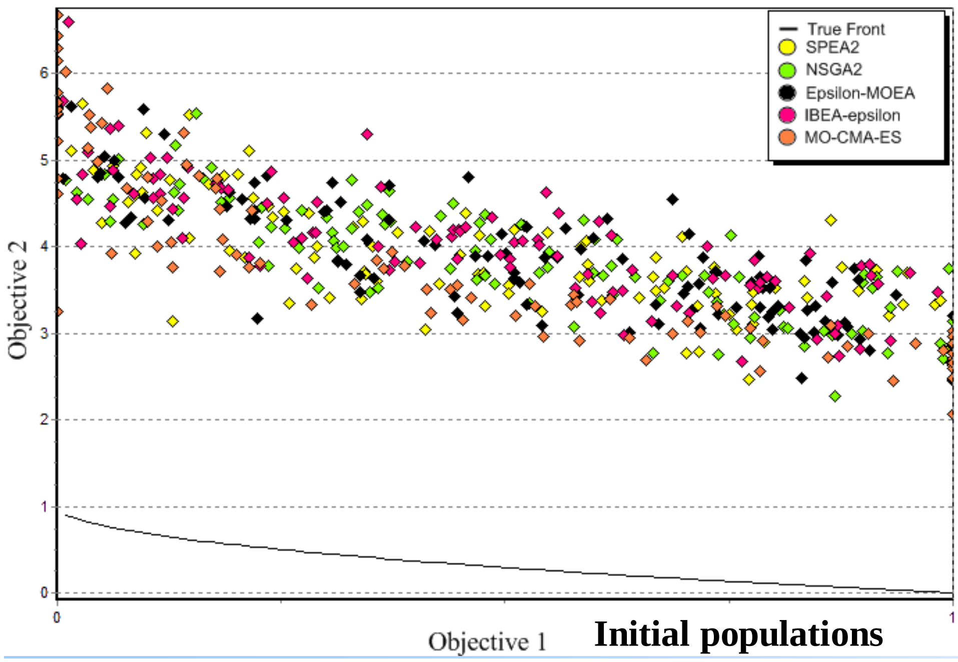
$$h_g: \quad \mathbb{R} \to \mathbb{R}_{\ge 0}, \quad x \mapsto \frac{x^2}{\lvert x \rvert + 0.1}$$

(Igel et al 2006)

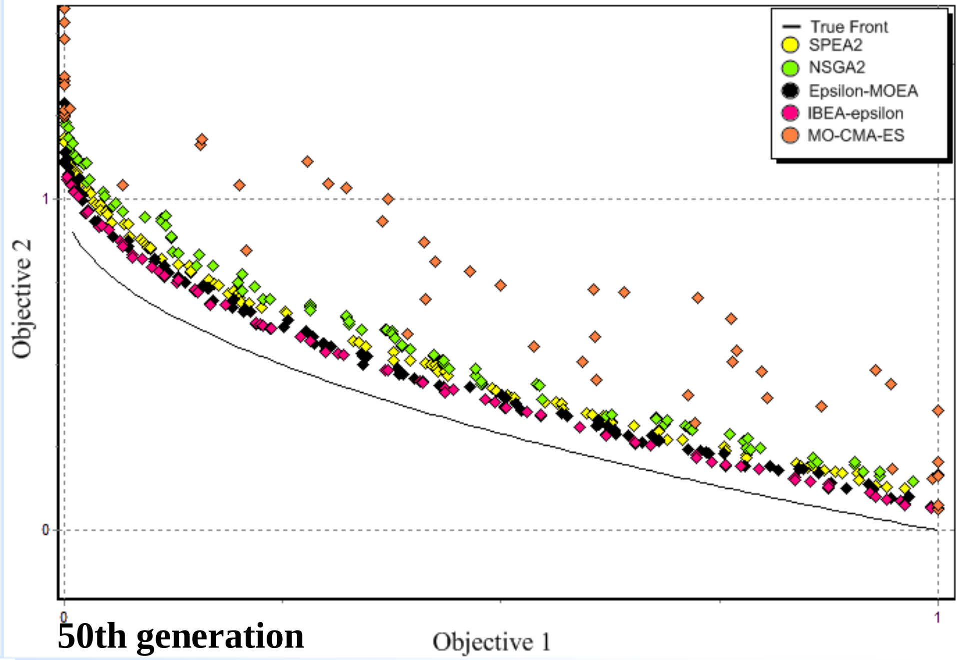
ZDT1 problem :



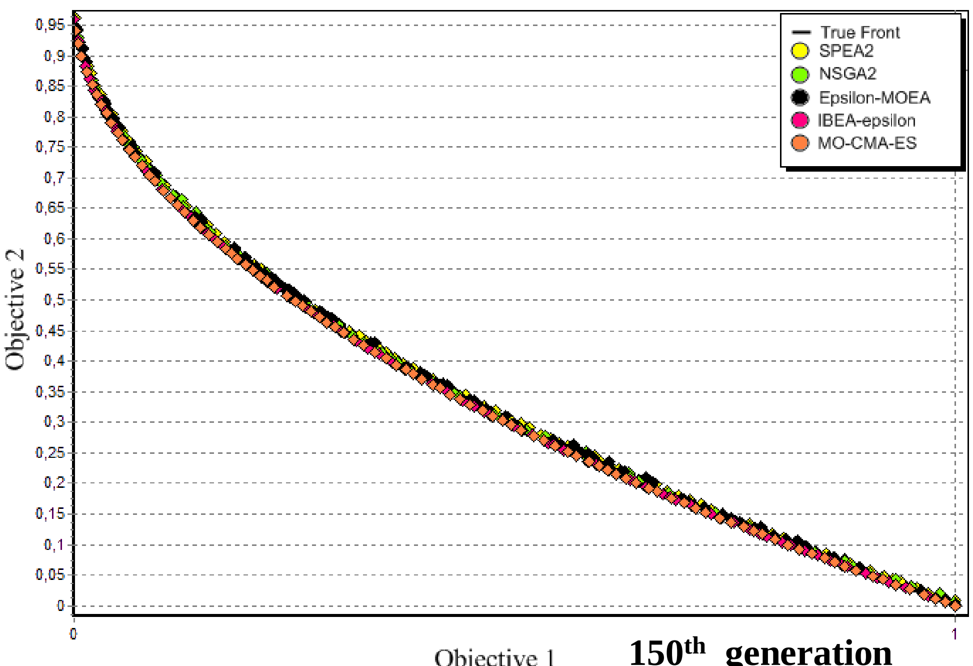
ZDT Problems ZDT1 convex



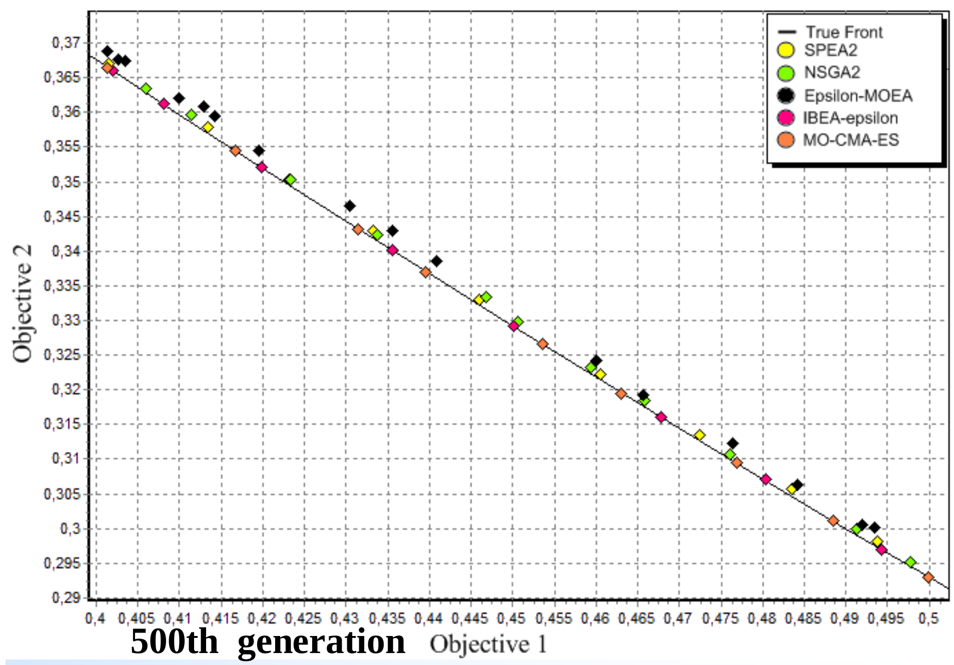
Initial populations



50th generation



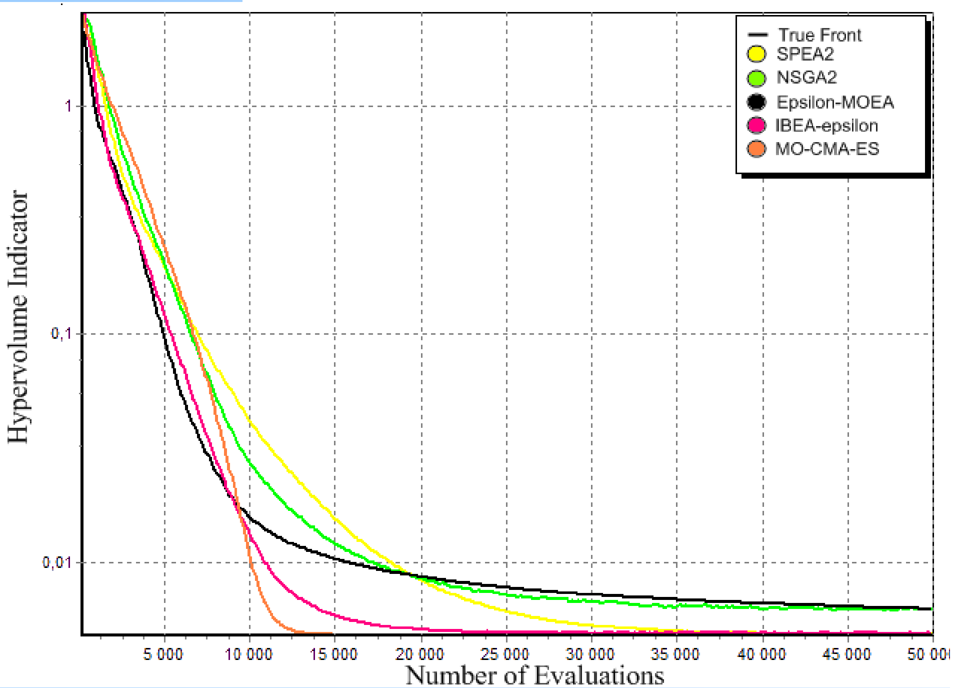
150th generation

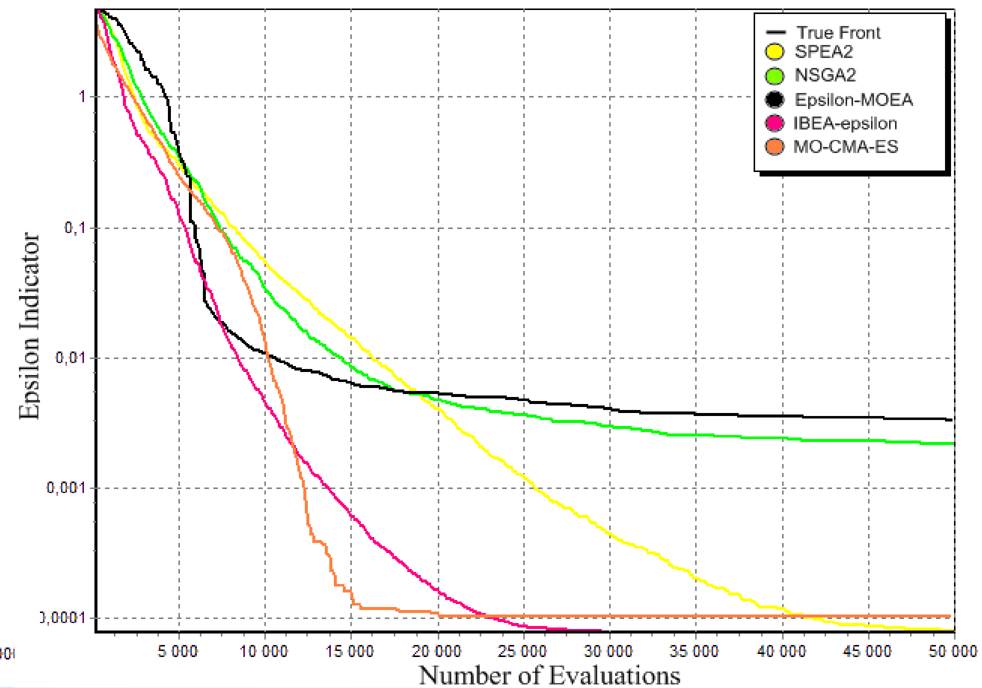


500th generation

ZDT1

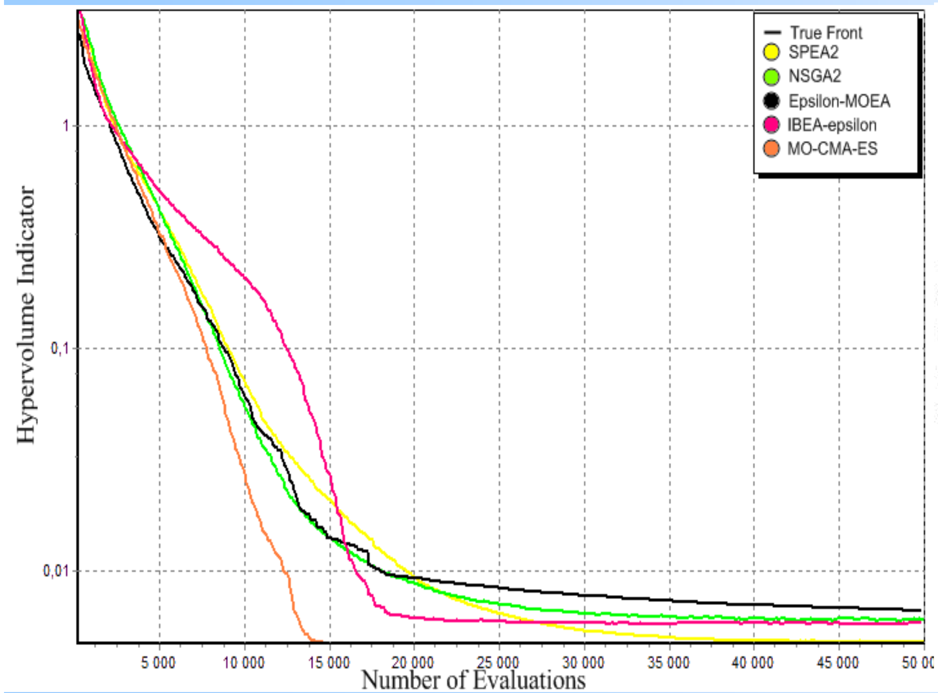
Number of Evaluations	Hypervolume indicator				
	I. NSGA2	II. ε-MOEA	III. SPEA2	IV. IBEA-ε	V. MO-CMA-ES
25000	0.00717 [II]	**0.00772**	0.00605 [I,II]	0.00490 [I,II,III]	0.00477 [I,II,III,IV]
50000	**0.00632**	**0.00622**	0.00482 [I,II]	0.00484 [I,II]	0.00475 [I,II,]
			ε - indicator		
25000	0.00360 [II]	**0.00473**	0.00123 [I,II]	0.00008 [I,II,III,V]	0.00010 [I,II,III]
50000	0.00216 [II]	0.00324	0.00008 [I,II]	0.00007 [I,II,III,V]	0.00010 [I,II,III]

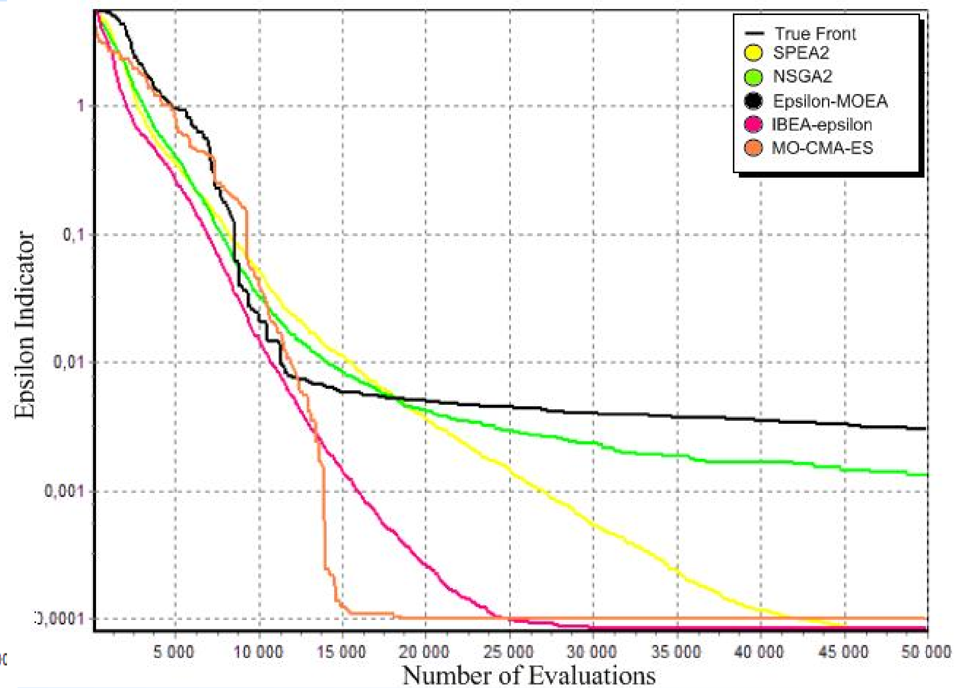


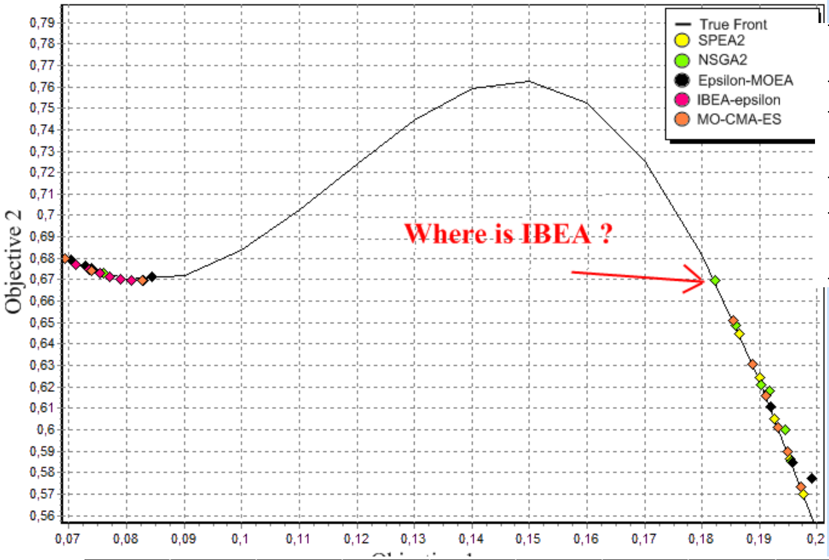


ZDT2 - concave

Number of Evaluations	Hypervolume indicator				
	I. NSGA2	II. ε-MOEA	III. SPEA2	IV. IBEA-ε	V. MO-CMA-ES
25000	0.00698 ^{II}	**0.00832**	0.00637 ^{I,II}	0.00587 ^{I,II,III}	0.00468 ^{I,II,III,IV}
50000	0.00609 ^{II}	**0.00658**	0.00474 ^{I,II,IV}	0.00577 ^{I,II}	0.00467 ^{I,II,IV}
	ε - indicator				
25000	0.00292 ^{II}	**0.00444**	0.00145 ^{I,II}	0.00010 ^{I,II,III}	0.00010 ^{I,II,III}
50000	0.00131 ^{II}	**0.00304**	0.00008 ^{I,II,V}	0.00008 ^{I,II,V}	0.00010 ^{I,II}

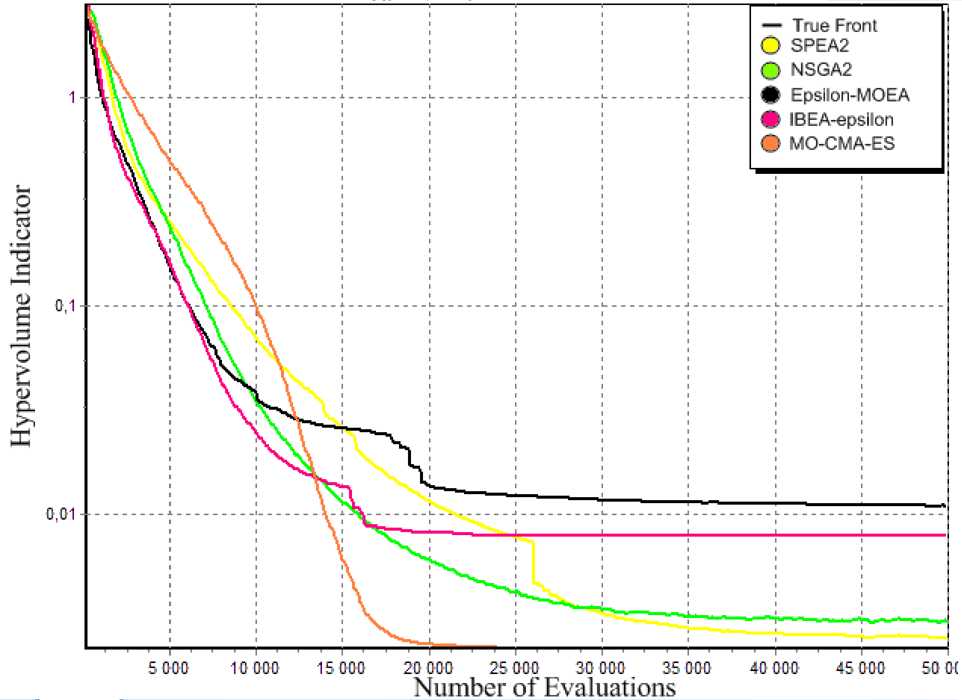


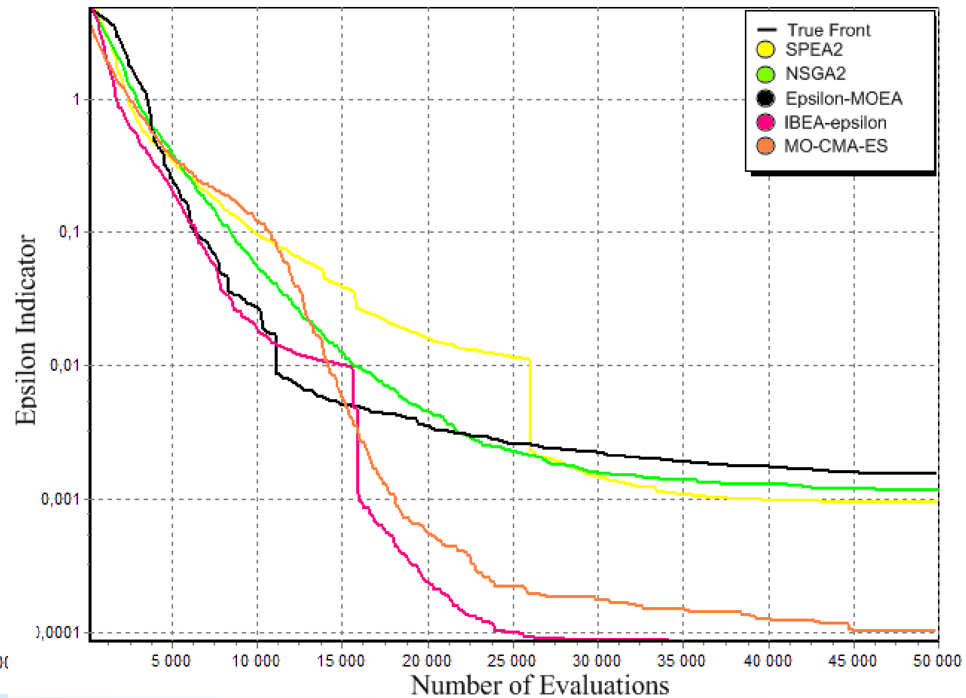




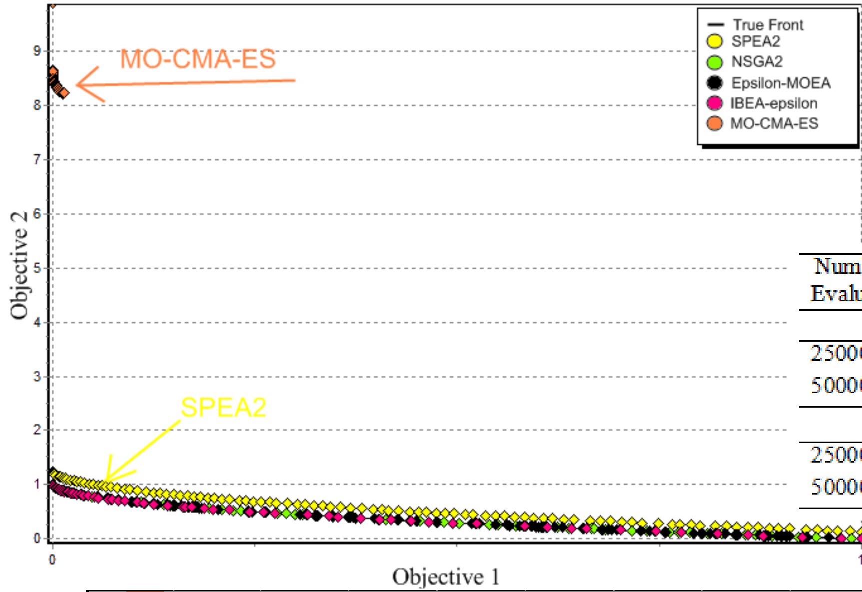
Number of Evaluations	Hypervolume indicator				
	I. NSGA2	II. ε -MOEA	III. SPEA2	IV. IBEA- ε	V. MO-CMA-ES
25000	0.00431 [II]	**0.01233**	0.00779 [I,II]	0.00800 [I,II,III]	0.00230 [I,II,III,IV]
50000	0.00307 [II,IV]	**0.01101**	0.00260 [I,II,IV]	0.00791 [I,II]	0.00228 [I,II,III,IV]
	ε - indicator				
25000	0.00227 [II,III]	**0.00260**	0.01143 [II]	0.00010 [I,II,III,V]	0.00022 [I,II,III]
50000	0.00115 [II]	**0.00151**	0.00094 [I,II]	0.00008 [I,II,III,V]	0.00010 [I,II,III]

ZDT3 - discontinuous

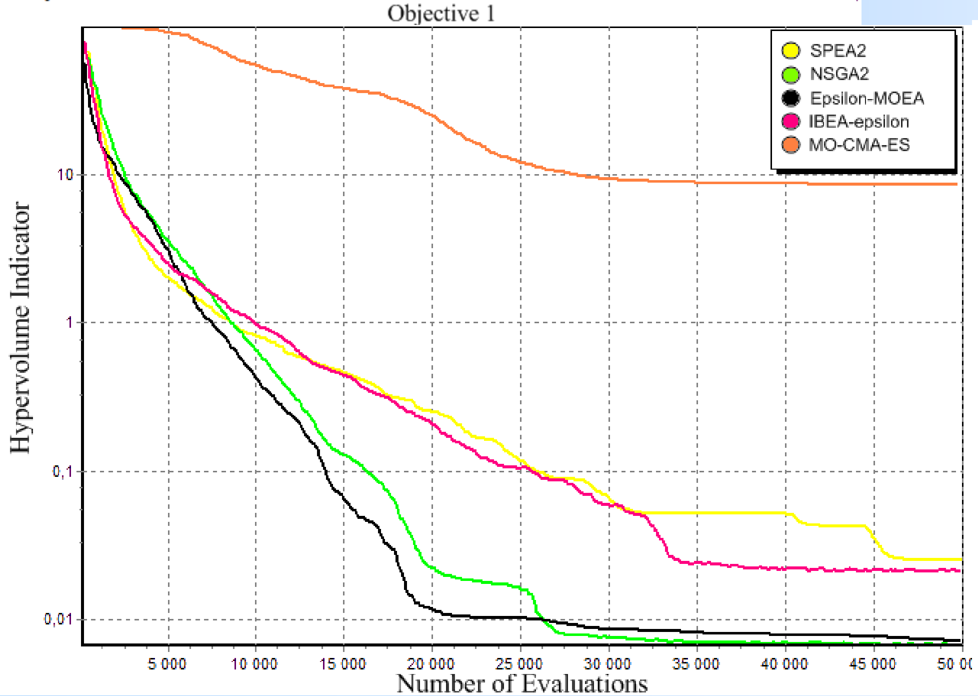


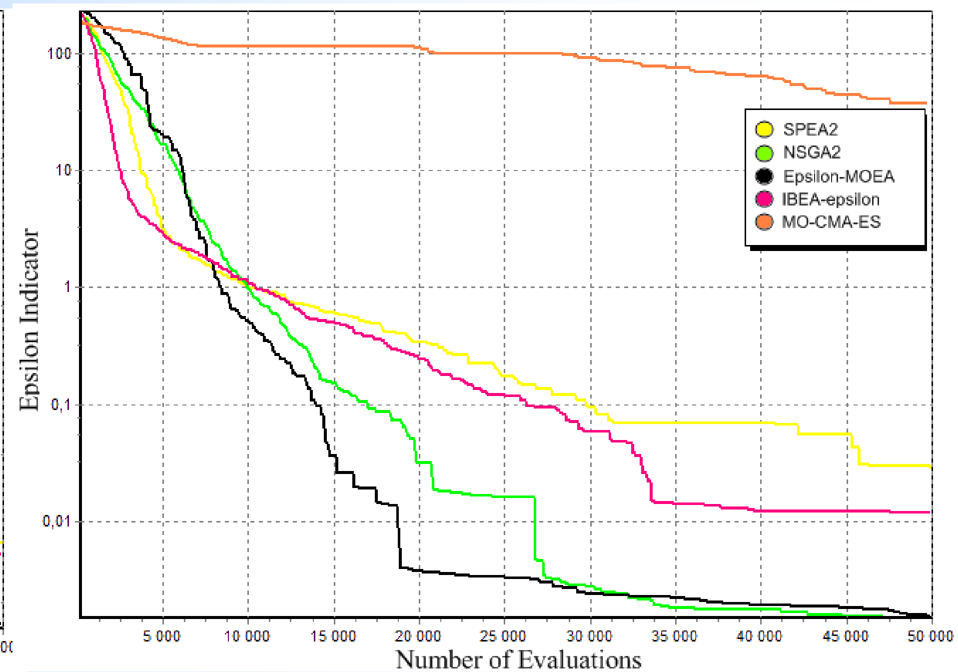


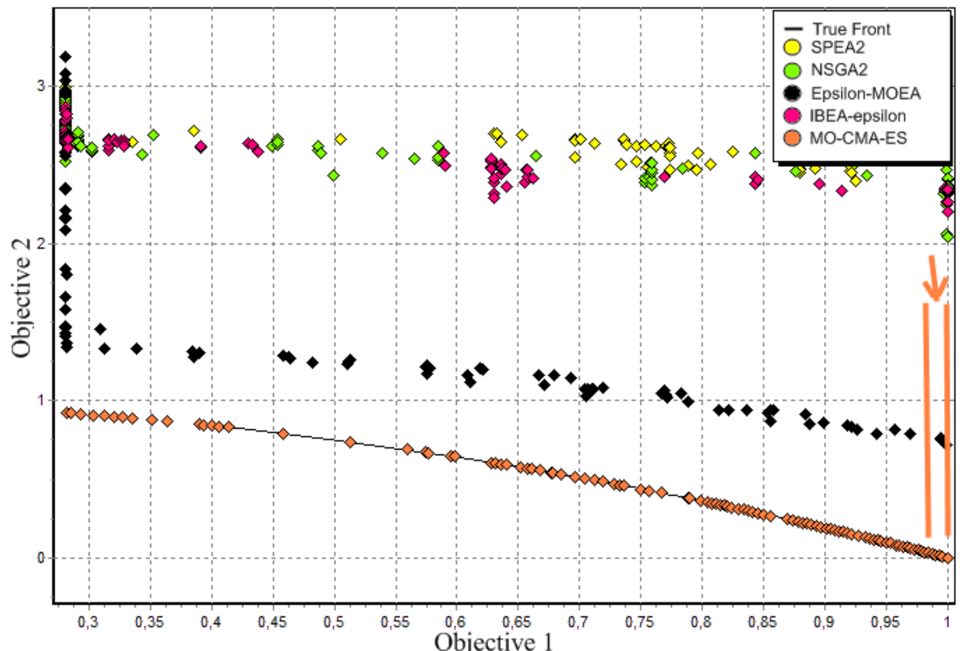
ZDT4 – separable problem with many local Pareto fronts



Number of Evaluations	Hypervolume indicator				
	I. NSGA2	II. ε -MOEA	III. SPEA2	IV. IBEA- ε	V. MO-CMA-ES
25000	0.01599 [III,IV,V]	0.01020 [I,III,IV,V]	0.11757 [V]	0.10536 [III,V]	**11.880**
50000	0.00684 [II,III,IV,V]	0.00722 [III,IV,V]	0.02501 [V]	0.02132 [III,V]	**8.3976**
	ε - indicator				
25000	0.01614 [III,IV,V]	0.00333 [I,III,IV,V]	0.17580 [V]	0.11921 [III,V]	**99.977**
50000	0.00144 [II,III,IV,V]	0.00155 [III,IV,V]	0.02941 [V]	0.01185 [III,V]	**36.914**

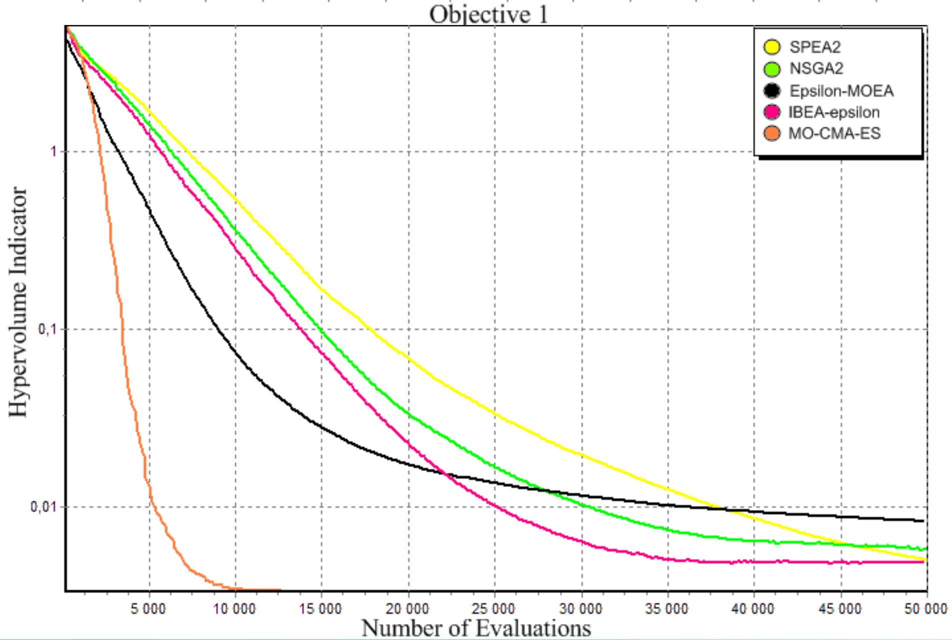


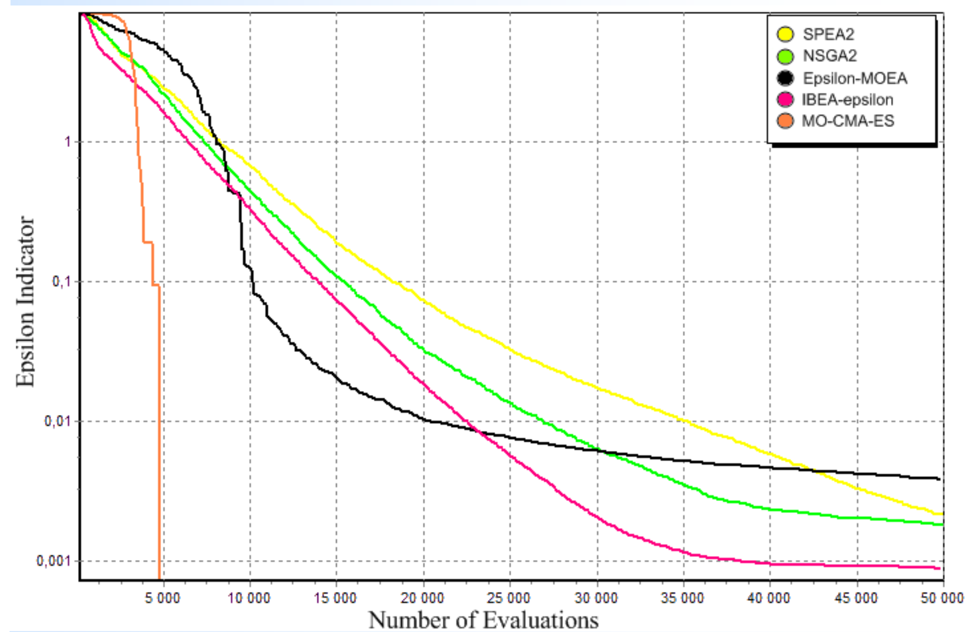




ZDT6 with non-uniformly distributed optimal solutions

Number of Evaluations	Hypervolume indicator				
	I. NSGA2	II. ε -MOEA	III. SPEA2	IV. IBEA-ε	V. MO-CMA-ES
25000	0.01682 $^{\mathrm{III}}$	0.01367 $^{\mathrm{I,III}}$	**0.03352**	0.01016 $^{\mathrm{I,II,III}}$	0.00336 $^{\mathrm{I,II,III,IV}}$
50000	0.00586 $^{\mathrm{II}}$	**0.00836**	0.00505 $^{\mathrm{I,II}}$	0.00487 $^{\mathrm{I,II,III}}$	0.00335 $^{\mathrm{I,II,III,IV}}$
			ε - indicator		
25000	0.01353 $^{\mathrm{III}}$	0.00759 $^{\mathrm{I,III}}$	**0.03286**	0.00568 $^{\mathrm{I,II,III}}$	0.00072 $^{\mathrm{I,II,III,IV}}$
50000	0.00182 $^{\mathrm{II,III}}$	**0.00385**	0.00214 $^{\mathrm{II}}$	0.00088 $^{\mathrm{I,II,III}}$	0.00072 $^{\mathrm{I,II,III,IV}}$



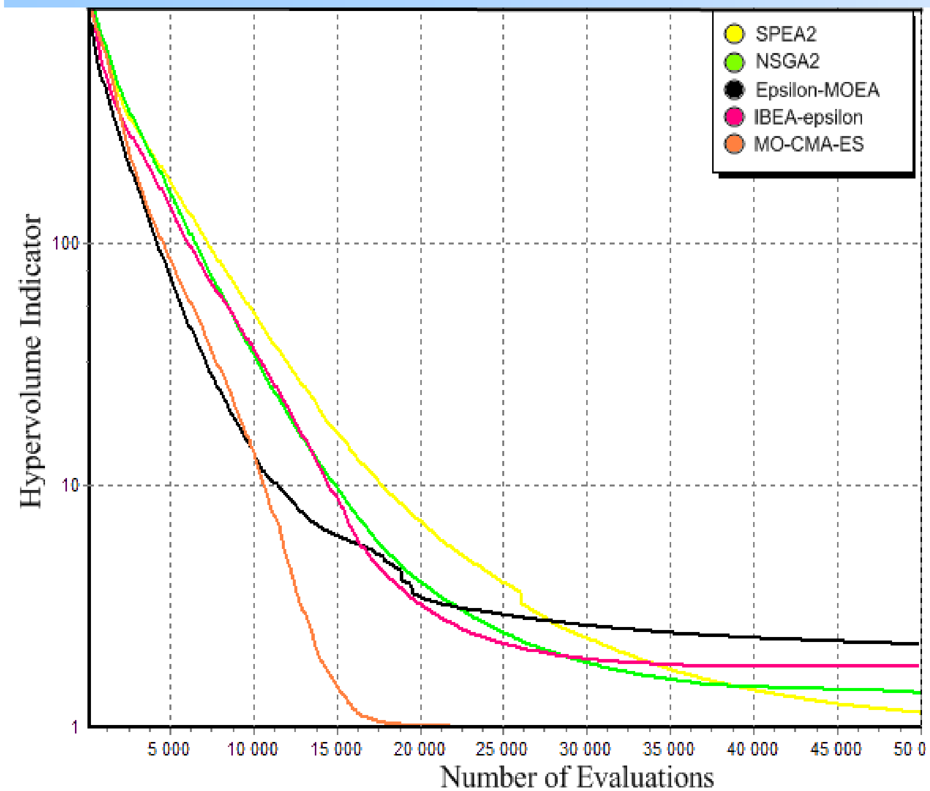


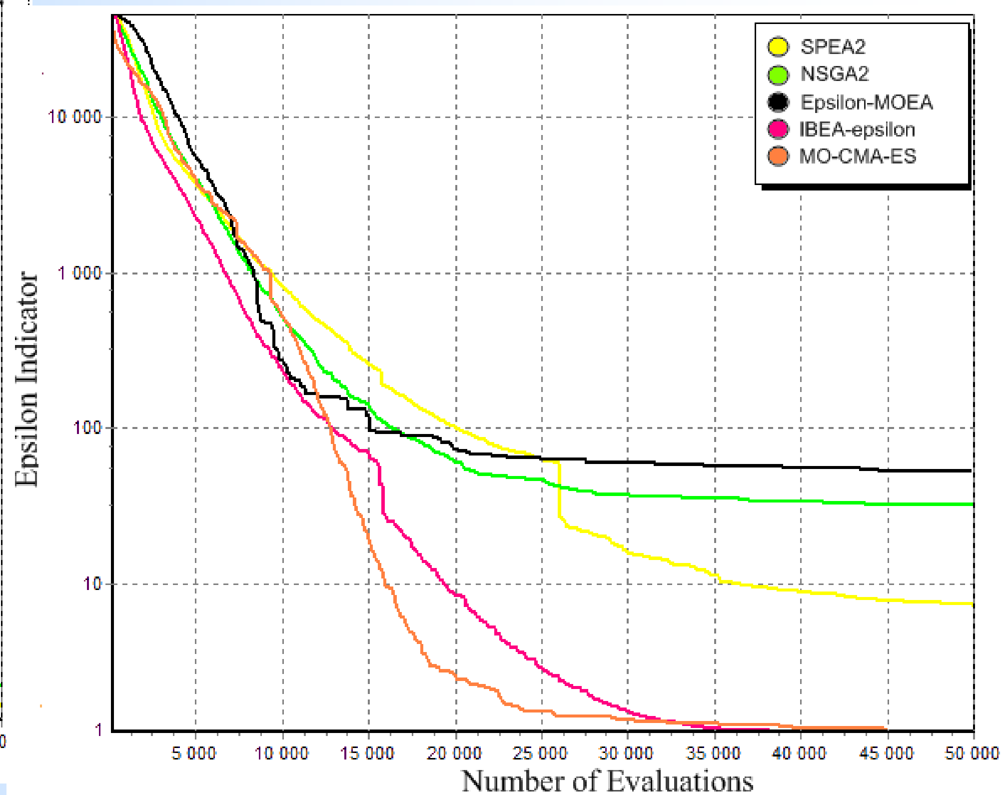
All ZDT problems

Algorithm	Hypervolume indicator 25000 evaluations				
	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
NSGA2	0.00717 ^{II}	0.00698 ^{II}	0.00431 ^{II}	0.01599 ^{III,IV,V}	0.01682 ^{III}
ε-MOEA	**0.00772**	**0.00832**	**0.01233**	0.01020 ^{I,III,IV,V}	0.01367 ^{I,III}
SPEA2	0.00605 ^{I,II}	0.00637 ^{I,II}	0.00779 ^{I,II}	0.11757 ^{V}	**0.03352**
IBEA-ε	0.00490 ^{I,II,III}	0.00587 ^{I,II,III}	0.00800 ^{I,II,III}	0.10536 ^{III,V}	0.01016 ^{I,II,III}
MO-CMA-ES	0.00477 ^{I,II,III,IV}	0.00468 ^{I,II,III,IV}	0.00230 ^{I,II,III,IV}	**11.880**	0.00336 ^{I,II,III,IV}
	50000 evaluations				
NSGA2	**0.00632**	0.00609 ^{II}	0.00307 ^{II,IV}	0.00684 ^{II,III,IV,V}	0.00586 ^{II}
ε-MOEA	**0.00622**	**0.00658**	**0.01101**	0.00722 ^{III,IV,V}	**0.00836**
SPEA2	0.00482 ^{I,II}	0.00474 ^{I,II,IV}	0.00260 ^{I,II,IV}	0.02501 ^{V}	0.00505 ^{I,II}
IBEA-ε	0.00484 ^{I,II}	0.00577 ^{I,II}	0.00791 ^{I,II}	0.02132 ^{III,V}	0.00487 ^{I,II,III}
MO-CMA-ES	0.00475 ^{I,II,}	0.00467 ^{I,II,IV}	0.00228 ^{I,II,III,IV}	**8.3976**	0.00335 ^{I,II,III,IV}

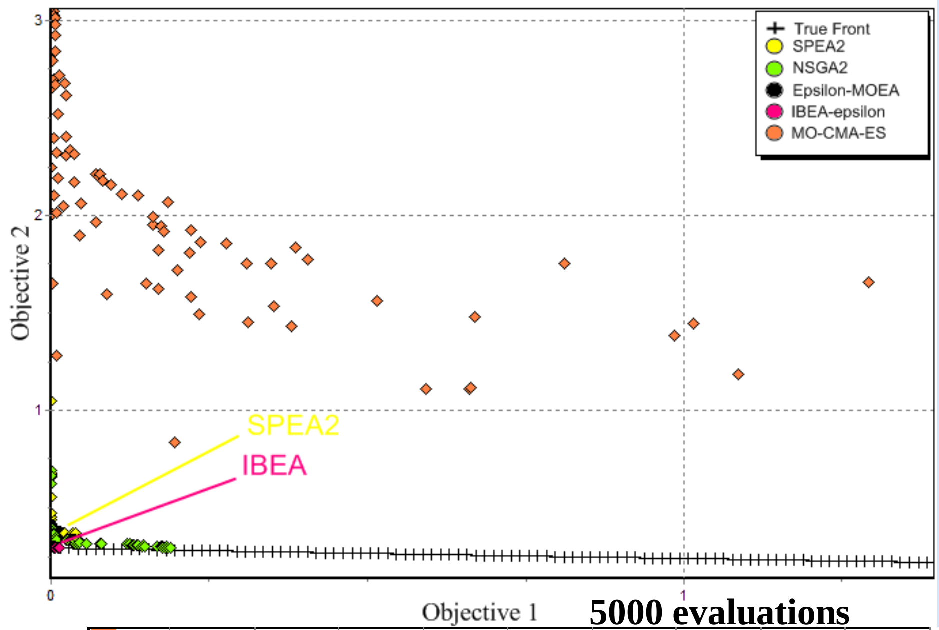
Algorithm	Epsilon indicator 25000 evaluations				
	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
NSGA2	0.00360 ^{II}	0.00292 ^{II}	0.00227 ^{II,III}	0.01614 ^{III,IV,V}	0.01353 ^{III}
ε-MOEA	**0.00473**	**0.00444**	**0.00260**	0.00333 ^{I,III,IV,V}	0.00759 ^{I,III}
SPEA2	0.00123 ^{I,II}	0.00145 ^{I,II}	0.01143 ^{II}	0.17580 ^{V}	**0.03286**
IBEA-ε	0.00008 ^{I,II,III,V}	0.00010 ^{I,II,III}	0.00010 ^{I,II,III,V}	0.11921 ^{III,V}	0.00568 ^{I,II,III}
MO-CMA-ES	0.00010 ^{I,II,III}	0.00010 ^{I,II,III}	0.00022 ^{I,II,III}	**99.977**	0.00072 ^{I,II,III,IV}
	50000 evaluations				
NSGA2	0.00216 ^{II}	0.00131 ^{II}	0.00115 ^{II}	0.00144 ^{II,III,IV,V}	0.00182 ^{II,III}
ε-MOEA	0.00324	**0.00304**	**0.00151**	0.00155 ^{III,IV,V}	**0.00385**
SPEA2	0.00008 ^{I,II}	0.00008 ^{I,II,V}	0.00094 ^{I,II}	0.02941 ^{V}	0.00214 ^{II}
IBEA-ε	0.00007 ^{I,II,III,V}	0.00008 ^{I,II,V}	0.00008 ^{I,II,III,V}	0.01185 ^{III,V}	0.00088 ^{I,II,III}
MO-CMA-ES	0.00010 ^{I,II,III}	0.00010 ^{I,II}	0.00010 ^{I,II,III}	**36.914**	0.00072 ^{I,II,III,IV}

All ZDT problems

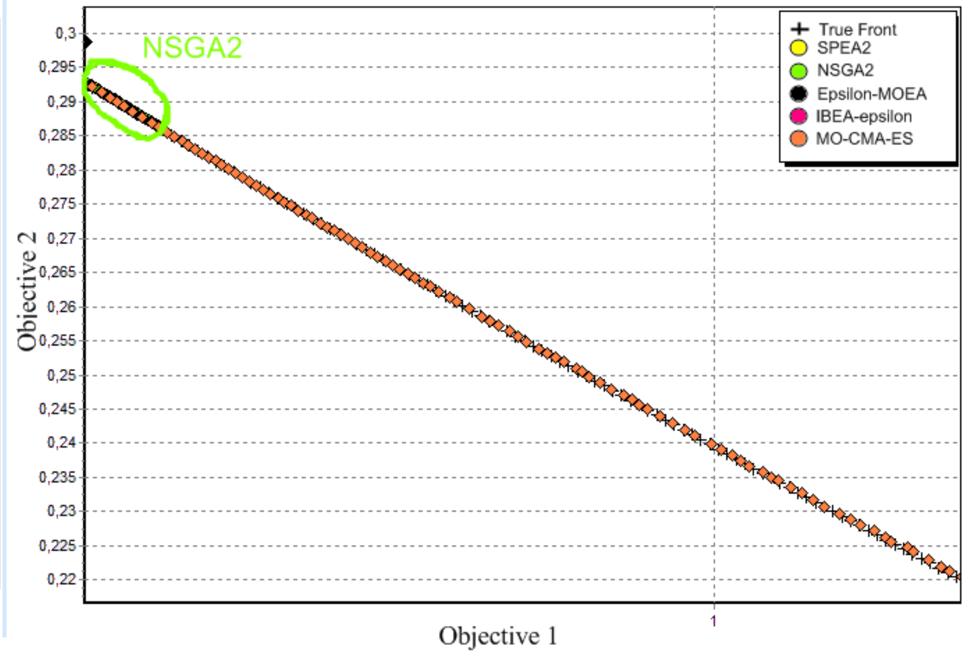


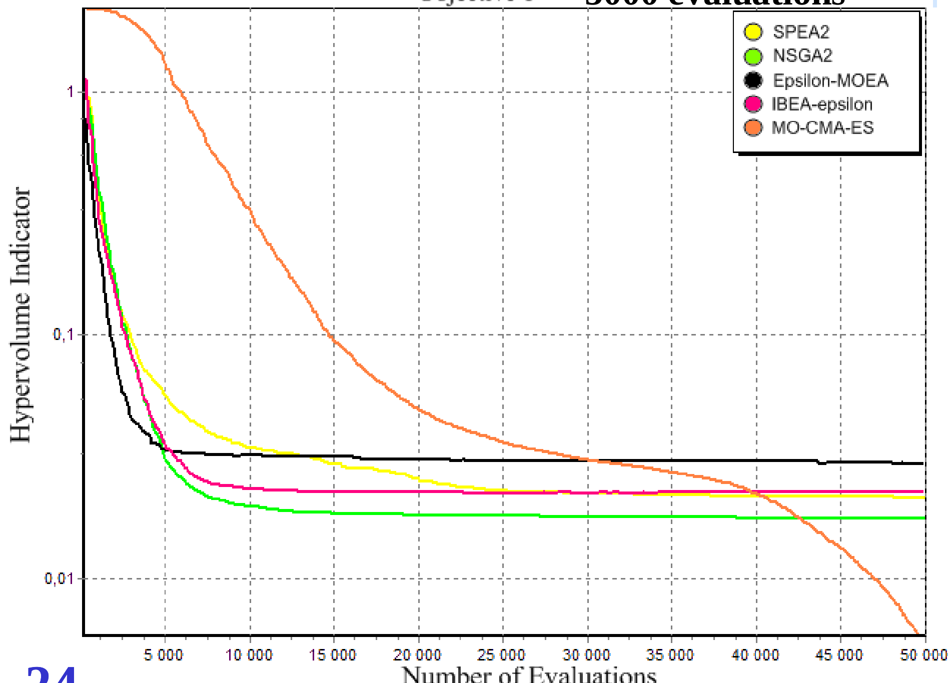


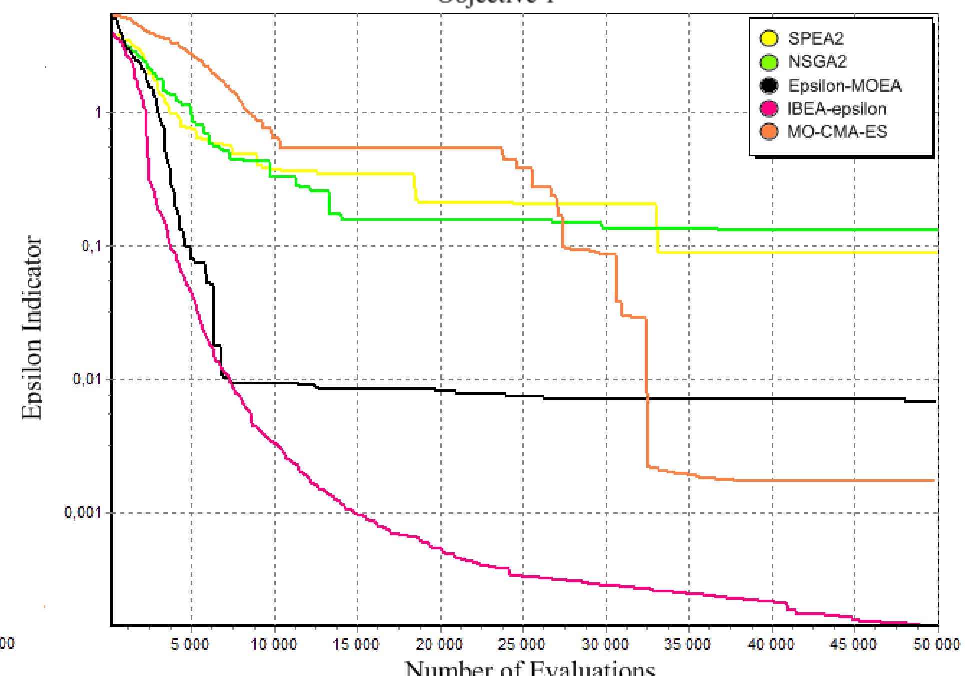
IHR Problems IHR1 convex



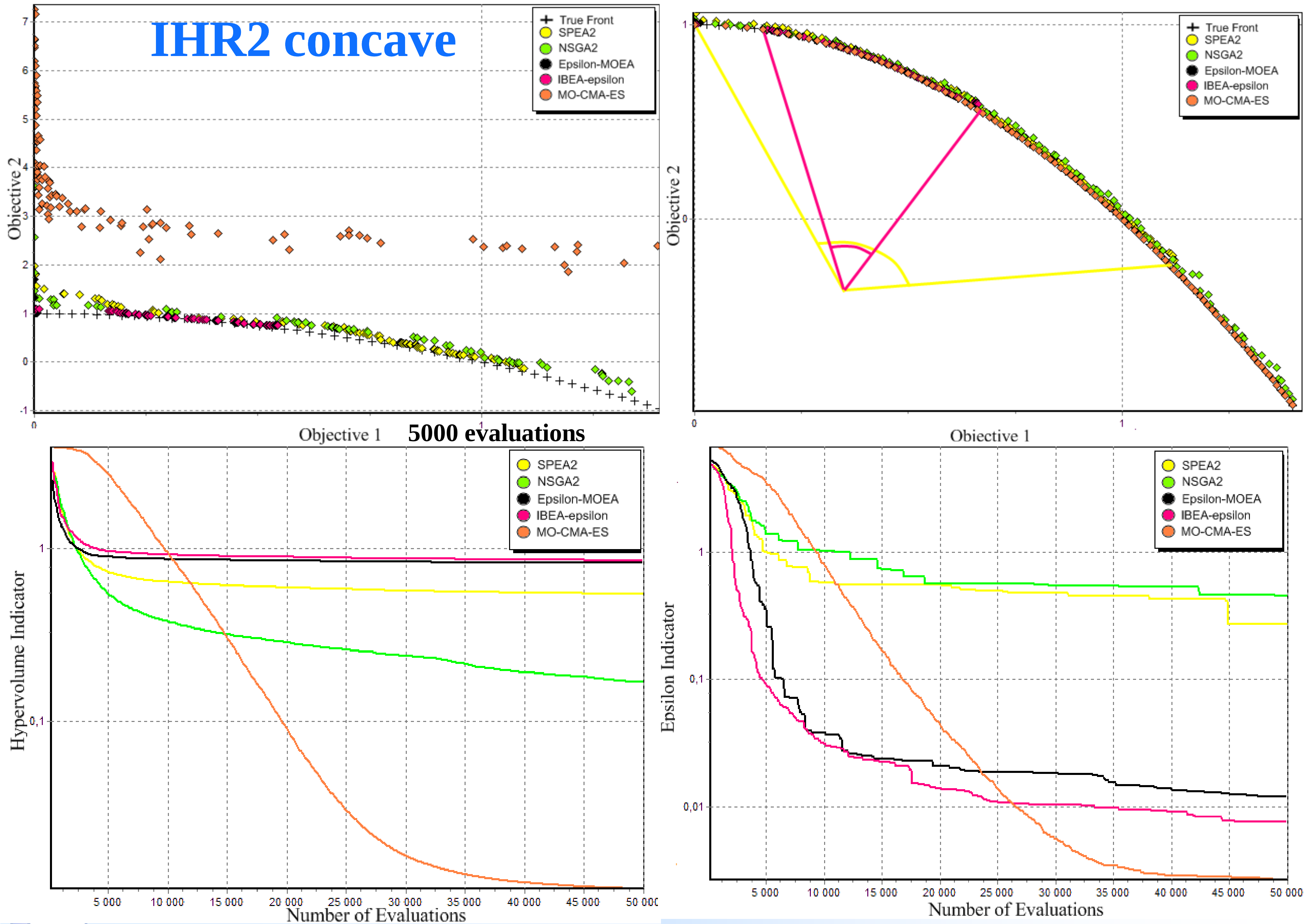
5000 evaluations



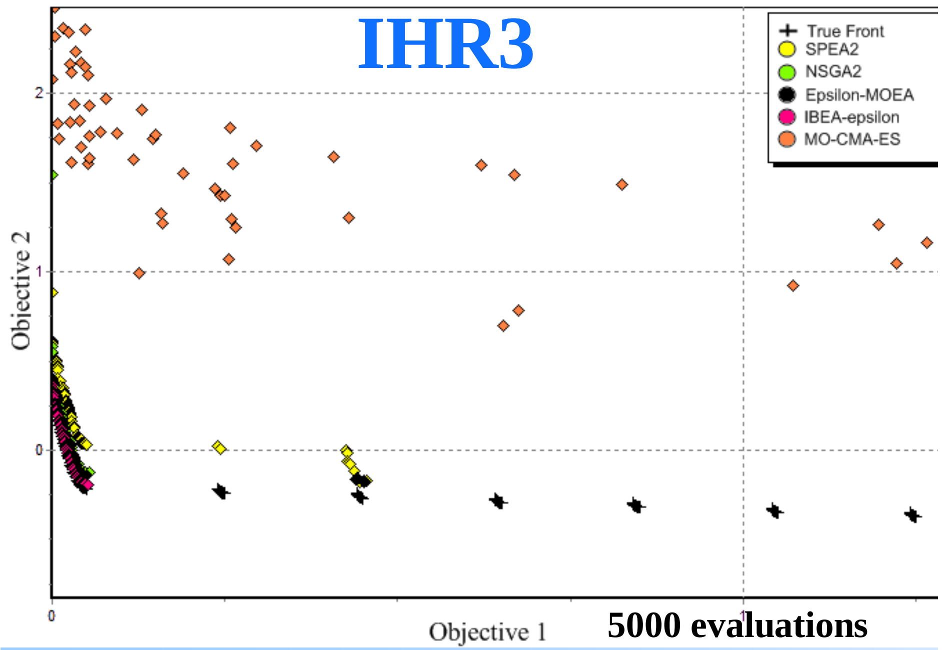




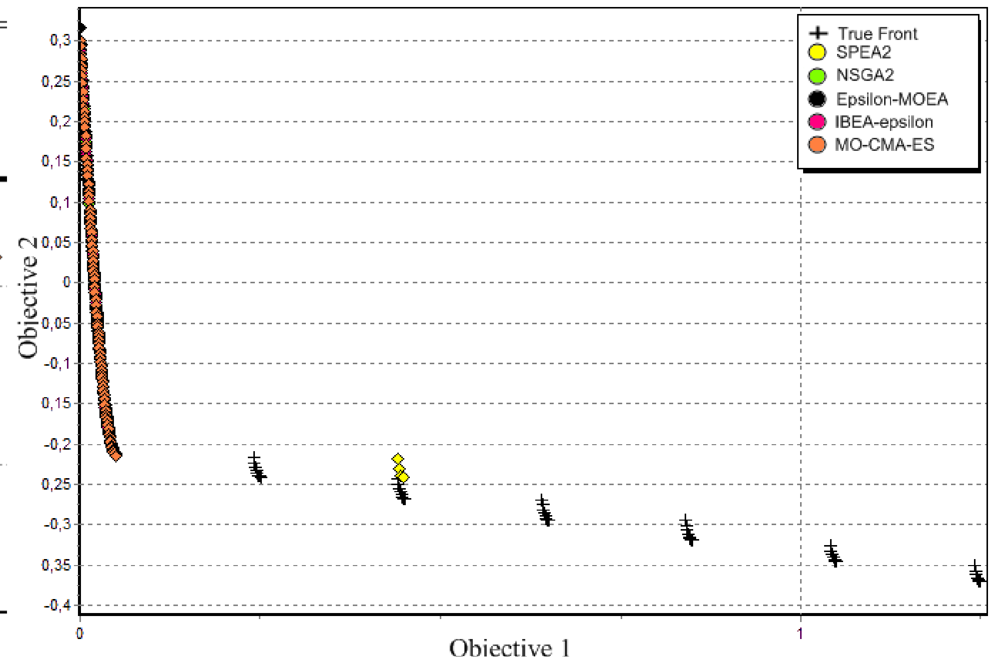
IHR2 concave

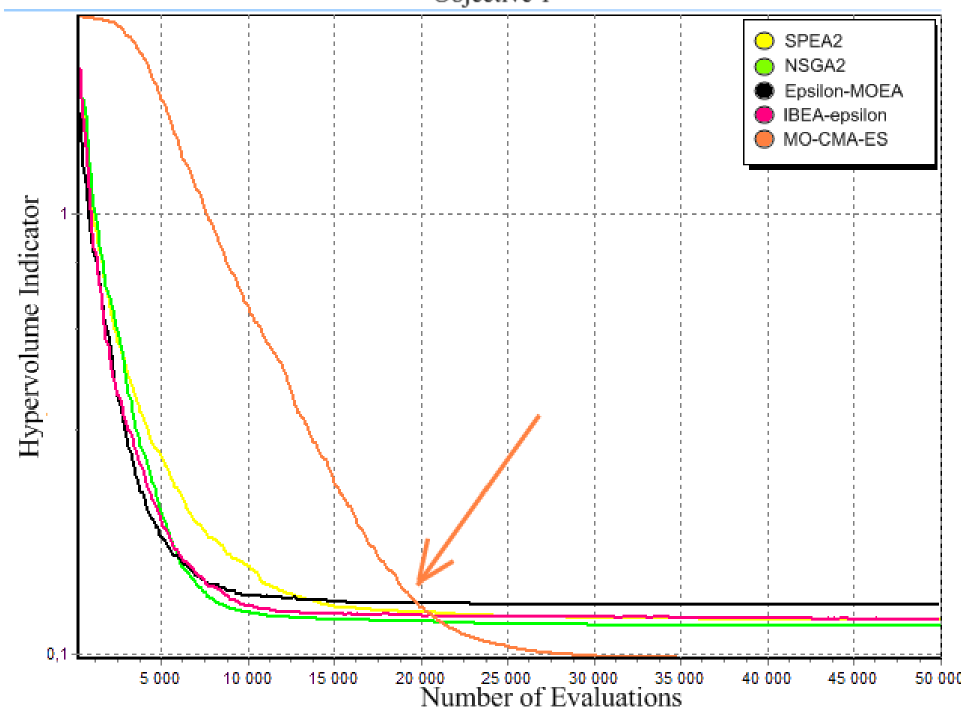


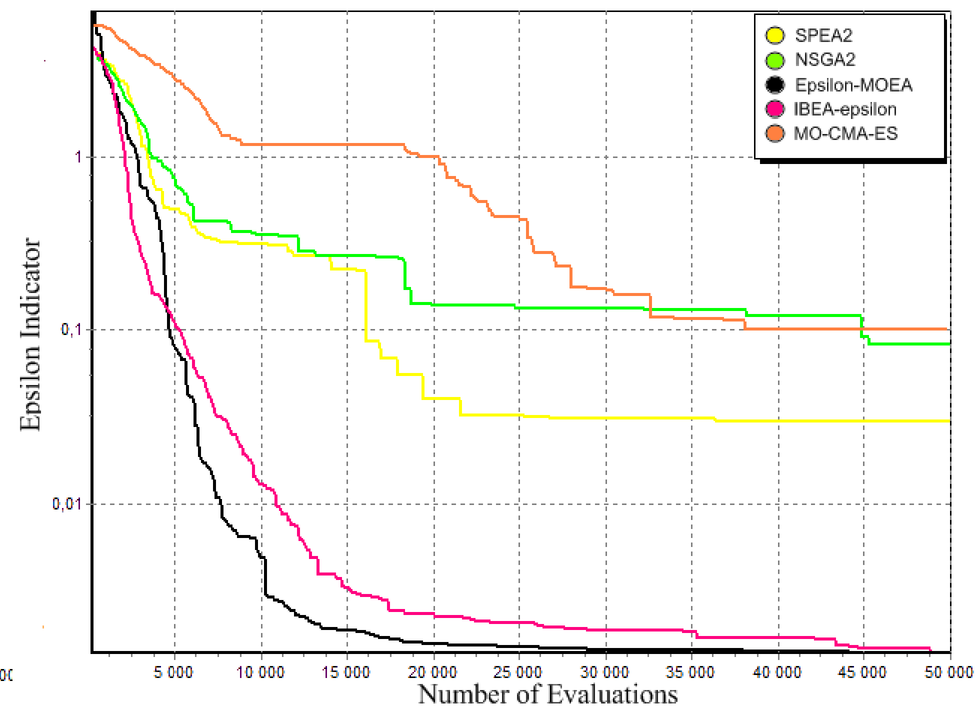
IHR3

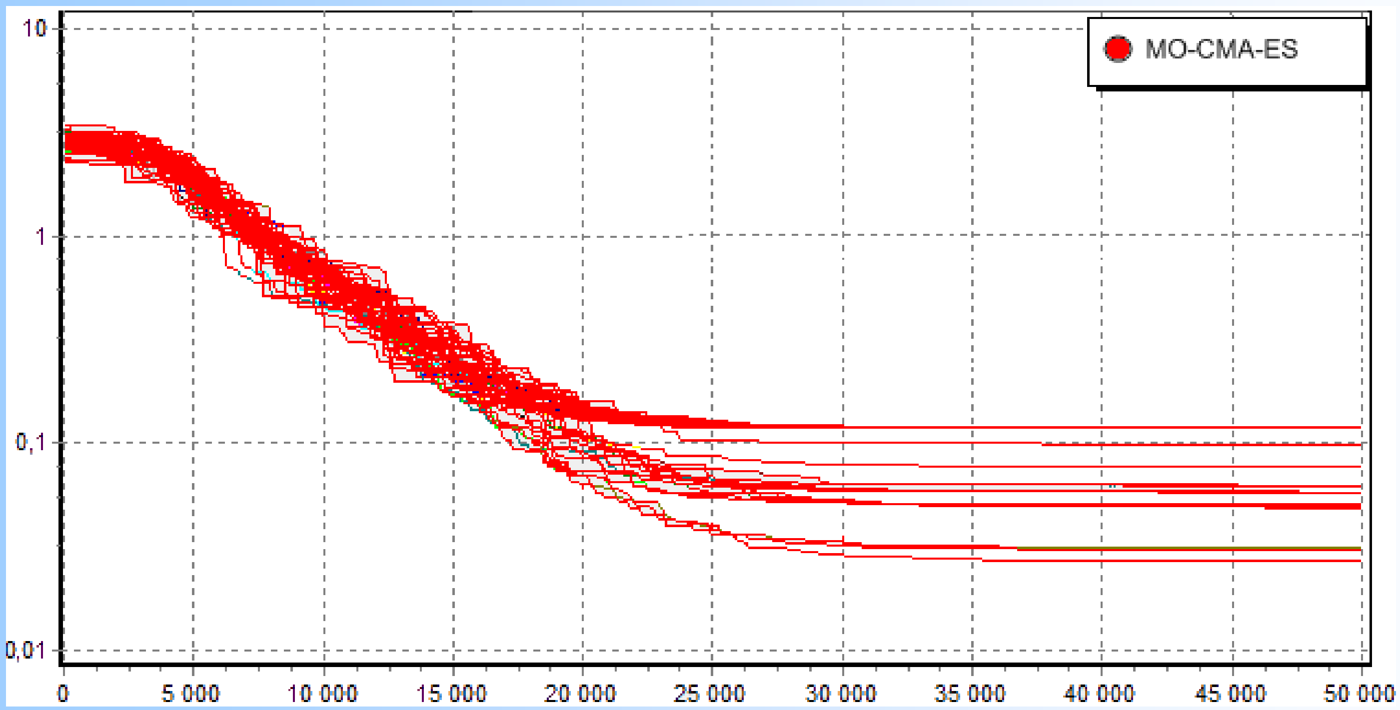


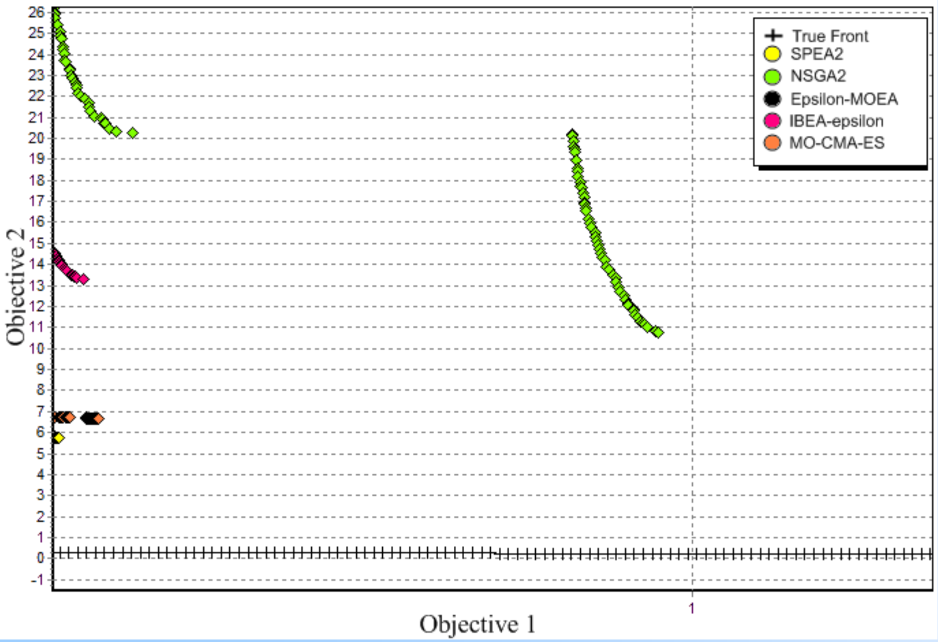
5000 evaluations



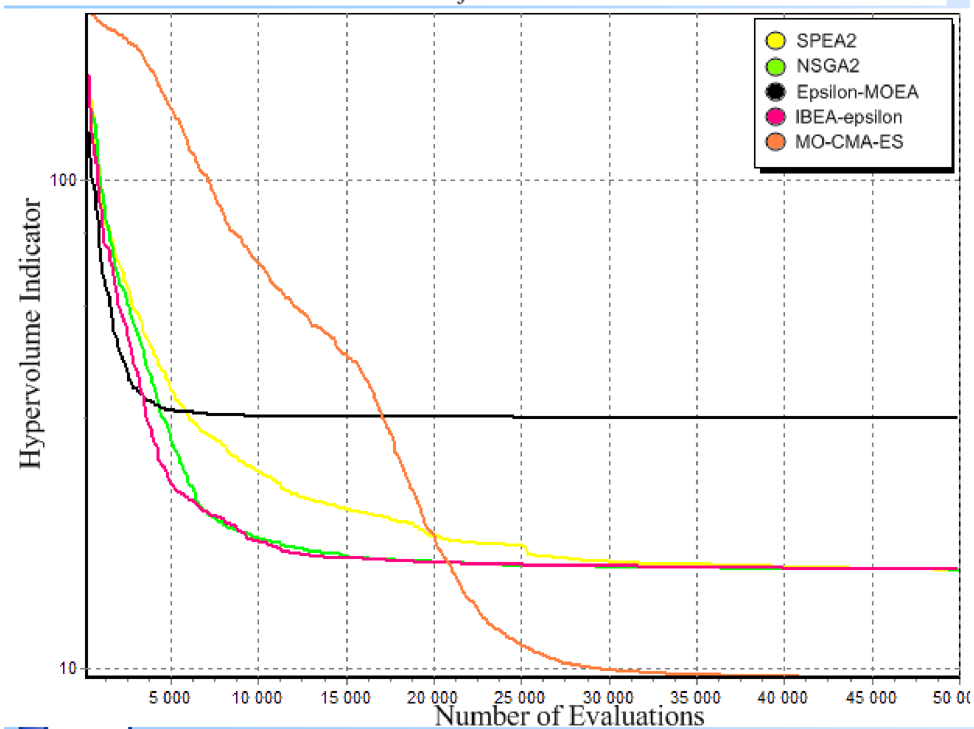


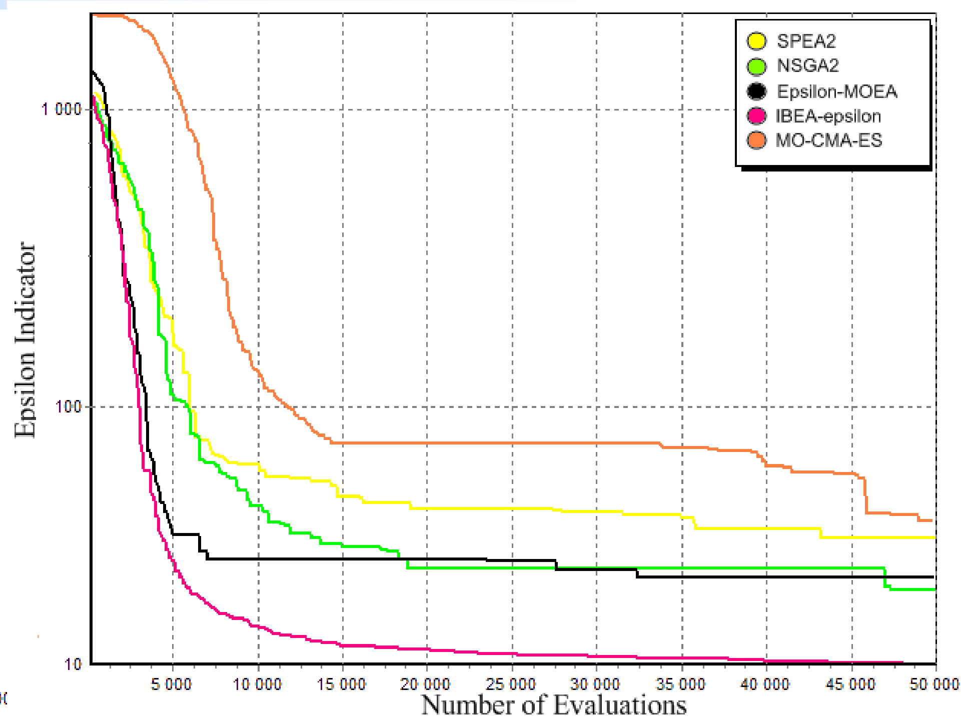


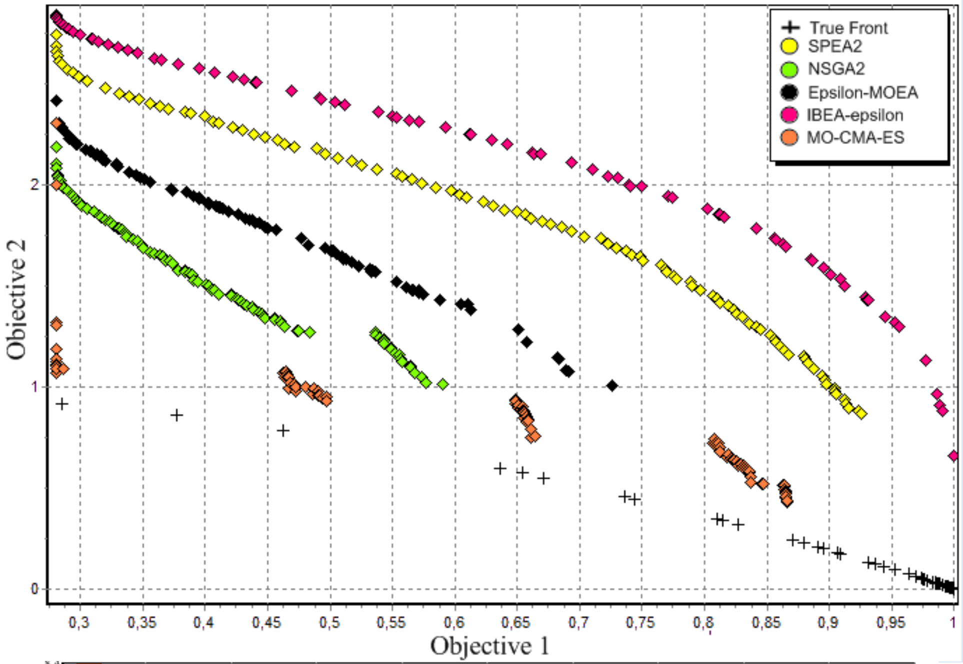

MO-CMA-ES
10
1
0,1
0,01
0
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000


True Front
SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
Objective 2
Objective 1

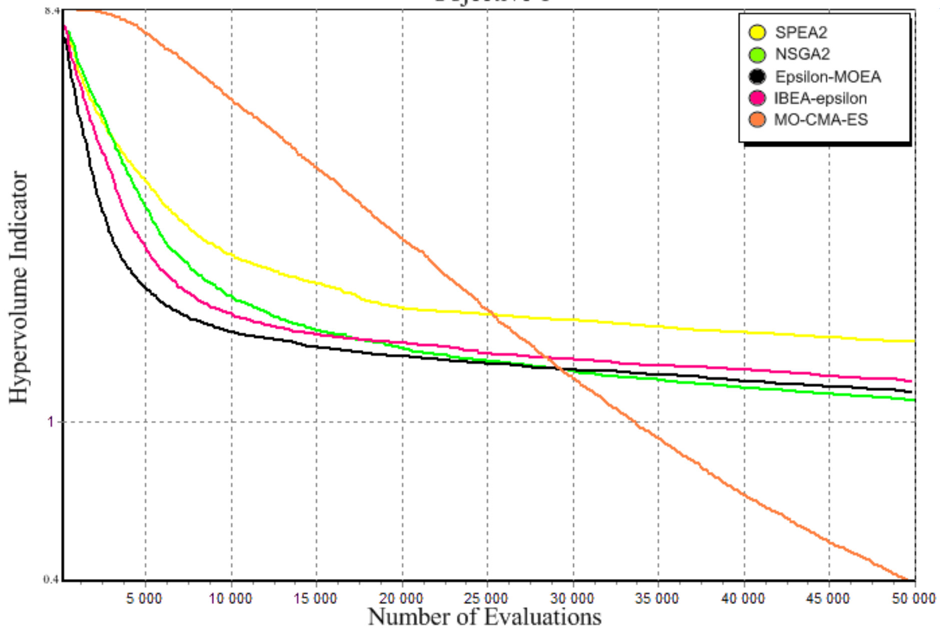
IHR4

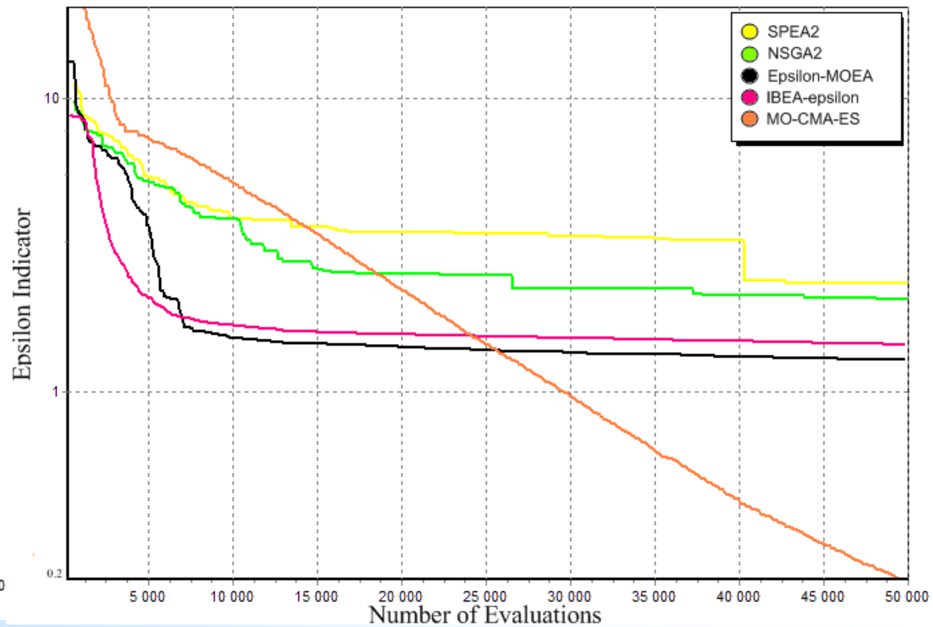

SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
Hypervolume Indicator
Number of Evaluations


SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
Epsilon Indicator
Number of Evaluations


True Front
SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
Objective 2
Objective 1

IHR6


SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
Hypervolume Indicator
Number of Evaluations

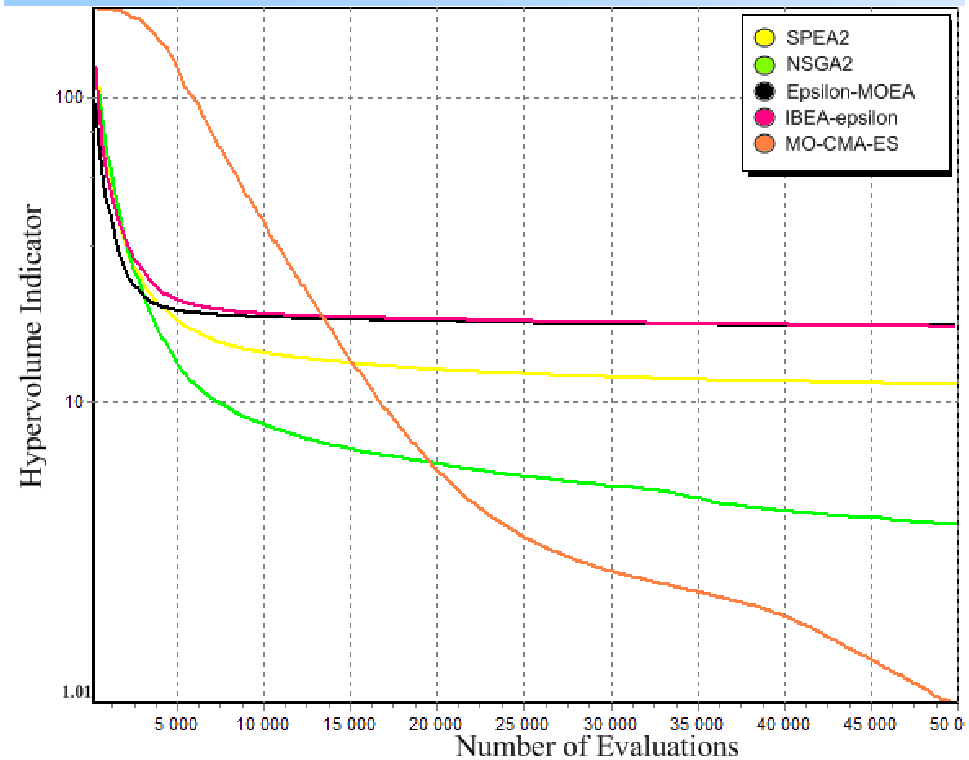

SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
Epsilon Indicator
Number of Evaluations

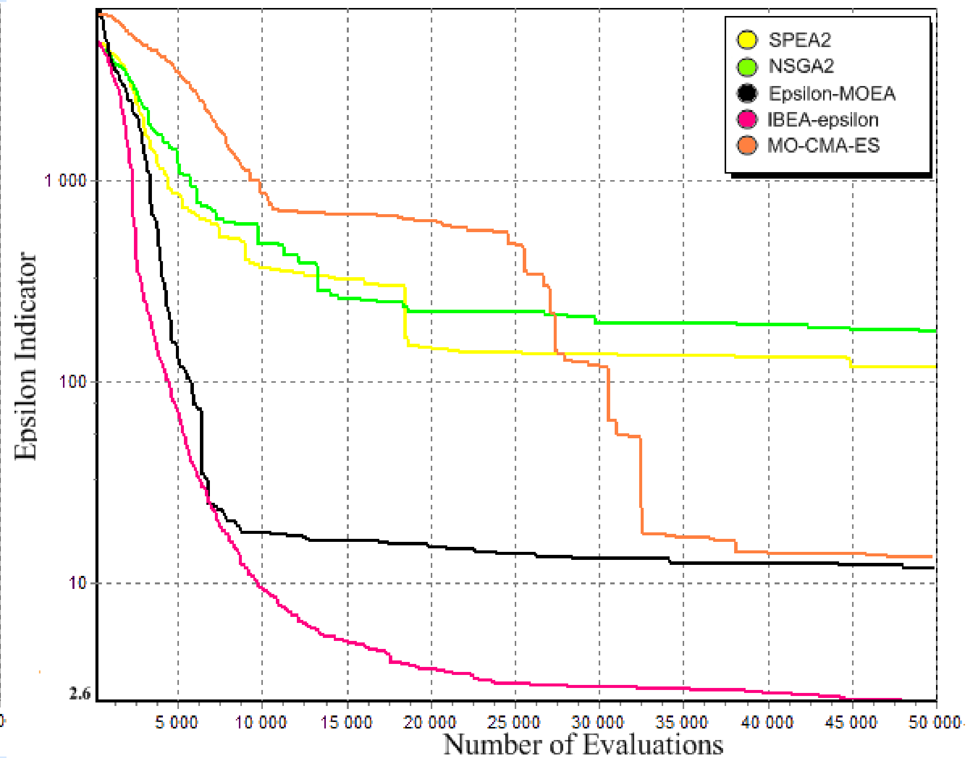
All IHR problems

Algorithm	Hypervolume indicator 25000 evaluations				
	IHR1	IHR2	IHR 3	IHR 4	IHR 6
NSGA2	0.0180 $^{\text{I,II,III,IV}}$	0.2599 $^{\text{II,III,IV}}$	0.1167 $^{\text{I,II,III}}$	16.283 $^{\text{II,III}}$	1.3690 $^{\text{III,IV,V}}$
ε -MOEA	0.0302 $^{\text{V}}$	0.8387 $^{\text{IV}}$	**0.1296**	**32.691**	1.3495 $^{\text{III,IV,V}}$
SPEA2	0.0228 $^{\text{II,V}}$	0.5840 $^{\text{II,IV}}$	0.1215 $^{\text{II}}$	17.837 $^{\text{II}}$	1.7380 $^{\text{V}}$
IBEA- ε	0.0224 $^{\text{II,V}}$	**0.8815**	0.1213 $^{\text{II}}$	16.343 $^{\text{II,III}}$	1.4208 $^{\text{III,V}}$
MO-CMA-ES	**0.0362**	0.0307 $^{\text{I,II,III,IV}}$	0.1037 $^{\text{I,II,III,IV}}$	11.164 $^{\text{I,II,III,IV}}$	**1.7834**
	50000 evaluations				
NSGA2	0.0176 $^{\text{II,III,IV}}$	0.1685 $^{\text{II,III}}$	0.1154 $^{\text{I,II,III}}$	15.966 $^{\text{II}}$	1.1125 $^{\text{II,III,IV}}$
ε -MOEA	**0.0295**	0.8222 $^{\text{IV}}$	**0.1291**	**32.561**	1.1654 $^{\text{III,IV}}$
SPEA2	0.0215 $^{\text{II,IV}}$	0.5467 $^{\text{II,IV}}$	0.1194 $^{\text{II}}$	15.918 $^{\text{II}}$	**1.5115**
IBEA- ε	0.0225 $^{\text{II}}$	**0.8502**	0.1195 $^{\text{II}}$	15.996 $^{\text{II}}$	1.2318 $^{\text{III}}$
MO-CMA-ES	0.0054 $^{\text{I,II,III,IV}}$	0.0138 $^{\text{I,II,III,IV}}$	0.0972 $^{\text{I,II,III,IV}}$	9.6045 $^{\text{I,II,III,IV}}$	0.4340 $^{\text{I,II,III,IV}}$

Algorithm	Epsilon indicator 25000 evaluations				
	IHR 1	IHR 2	IHR 3	IHR 4	IHR 6
NSGA2	0.1558 $^{\text{III,V}}$	**0.5595**	0.1334 $^{\text{V}}$	28.425 $^{\text{II,III,V}}$	2.4908 $^{\text{III}}$
ε -MOEA	0.0074 $^{\text{I,III,V}}$	0.0184 $^{\text{I,III}}$	0.0015 $^{\text{II,III,IV}}$	30.060 $^{\text{III,V}}$	1.3827 $^{\text{II,III,IV,V}}$
SPEA2	0.2084 $^{\text{V}}$	0.4951 $^{\text{I}}$	0.0326 $^{\text{I,V}}$	45.062 $^{\text{V}}$	**3.4461**
IBEA- ε	0.0003 $^{\text{I,II,III,V}}$	0.0109 $^{\text{I,II,III}}$	0.0020 $^{\text{I,III,V}}$	14.594 $^{\text{I,II,III,V}}$	1.5422 $^{\text{I,III}}$
MO-CMA-ES	**0.3810**	0.0107 $^{\text{I,II,III}}$	**0.4483**	**74.993**	1.4498 $^{\text{I,III,IV}}$
	50000 evaluations				
NSGA2	**0.1329**	**0.4494**	0.0819 $^{\text{V}}$	24.148 $^{\text{II,III,V}}$	2.0606 $^{\text{III}}$
ε -MOEA	0.0634 $^{\text{I,III}}$	0.0118 $^{\text{I,III}}$	0.0013 $^{\text{I,III,V}}$	26.500 $^{\text{III,V}}$	1.2779 $^{\text{I,III,IV}}$
SPEA2	0.0882 $^{\text{I}}$	0.2684 $^{\text{I}}$	0.0298 $^{\text{I,V}}$	36.220 $^{\text{V}}$	**2.3494**
IBEA- ε	0.0001 $^{\text{I,II,III,V}}$	0.0075 $^{\text{I,II,III}}$	0.0014 $^{\text{I,III,V}}$	13.470 $^{\text{I,II,III,V}}$	1.4465 $^{\text{I,III}}$
MO-CMA-ES	0.0017 $^{\text{I,II,III}}$	0.0026 $^{\text{I,II,III,IV}}$	**0.1002**	**39.386**	0.2236 $^{\text{I,II,III,IV}}$

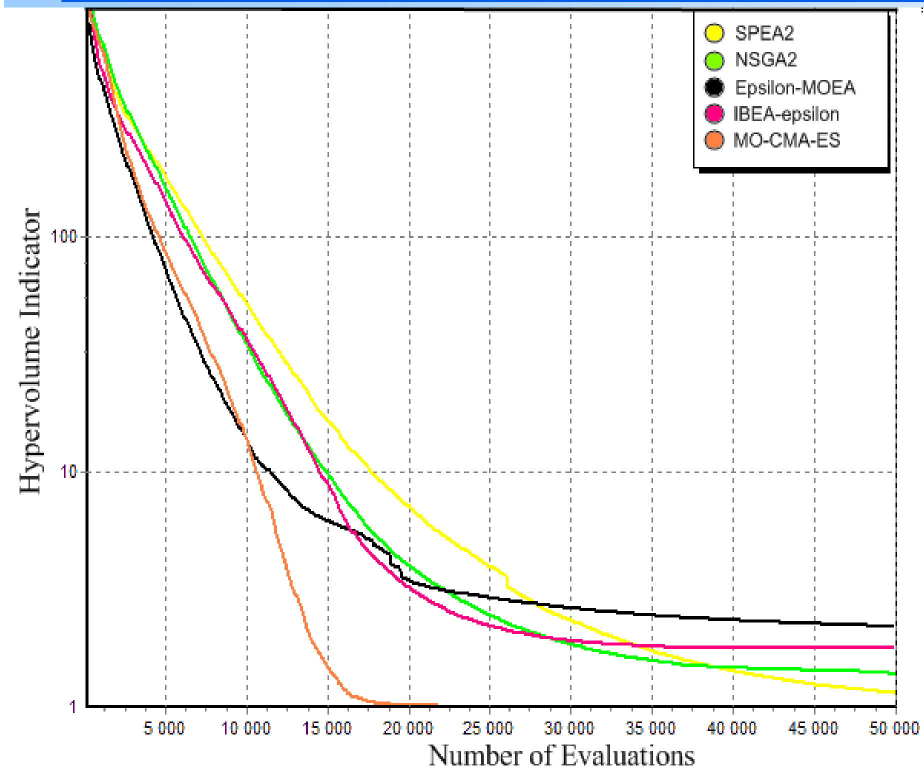
All IHR problems



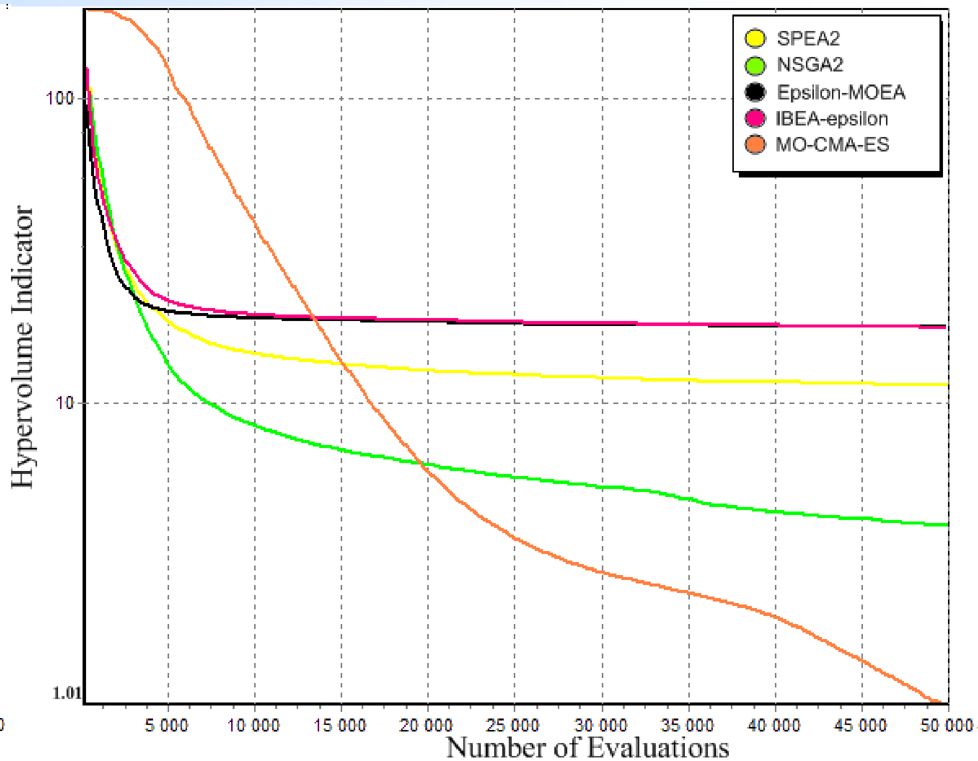


Summary

1.The MO-CMA-ES outperforms all other algorithms.
2.The NSGA-2 has the second rank.
3.The CMA-ES outperforms SBX+Polynomial mutation on rotated IHR problems.
4.The IBEA-epsilon worse than SPEA2.
5.The IBEA-epsilon and Epsilon-MOEA are the worst algorithms on IHR problems.



ZDT



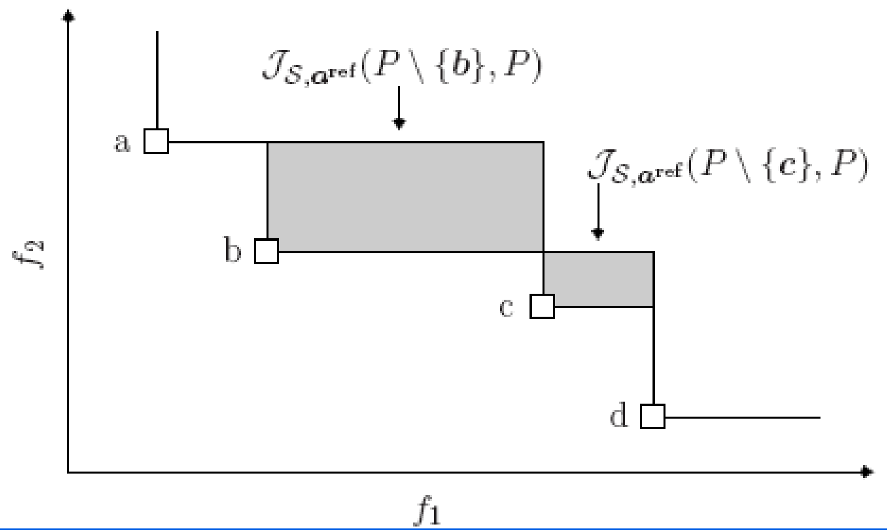
IHR

Nondominated Sorting Genetic Algorithm 2 based on hypervolume indicator as the second level sorting criterion (*S*-NSGA2).

The original NSGA2 use the crowding-distance as second level sorting criterion, while other criterions can be applied.

The *S-NSGA -2* based on the contribution hypervolume (*S -measure)* as second criterion introduced by Voss (2007) is the more contemporary version of original algorithm.

In original NSGA2 if a front cannot fit into archive entirely, the individuals from t his from further ranked according to the crowding distance metric. The Hypervolume and Epsilon indicators can be also applied instead of crowding distance metric.



Example of the hypervolume-indicator for a set *P={a,b,c,d}*. The objective vectors *b* and *c* are evaluated. In advance, the boundary elements *a* and *d* have been assigned a contribution rank |*P*|=*4* and |P|-1=3 .

Let F is the front which cannot fit into archive A entirely, because in A there are only N free positions for new individuals. Then $K=|F|-N$ is the number of individuals in F which should be deleted.

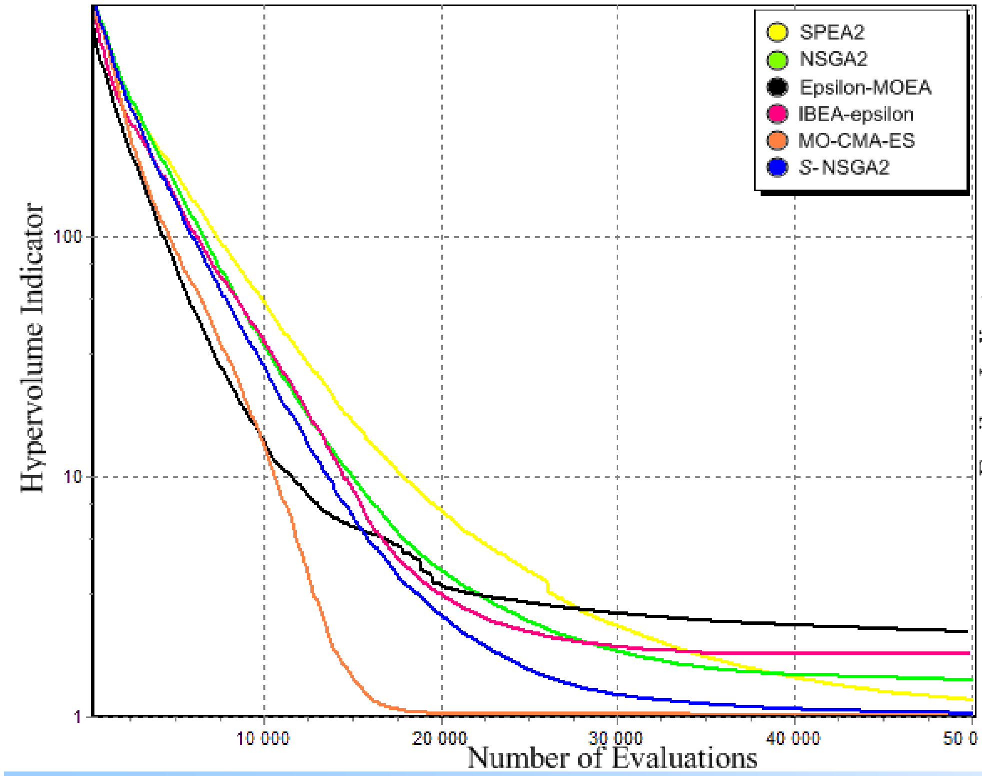
The procedure of ranking consists of K iterations. At each iteration the worst individual (in sense contributing hypervolume) should be deleted.

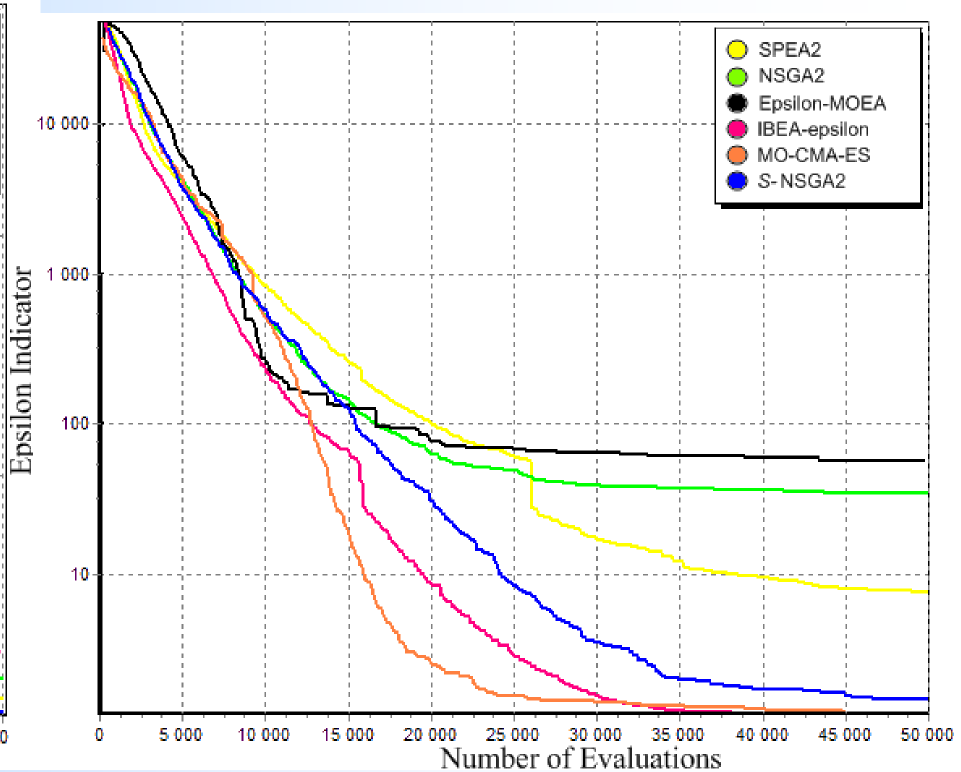
The worst individual a is the individual which has the minimum contribution hypervolume, i.e. the portion of objective space exclusively weakly dominated by a.

After K iterations the size of F reduces to N and then front F can fit into entirely.

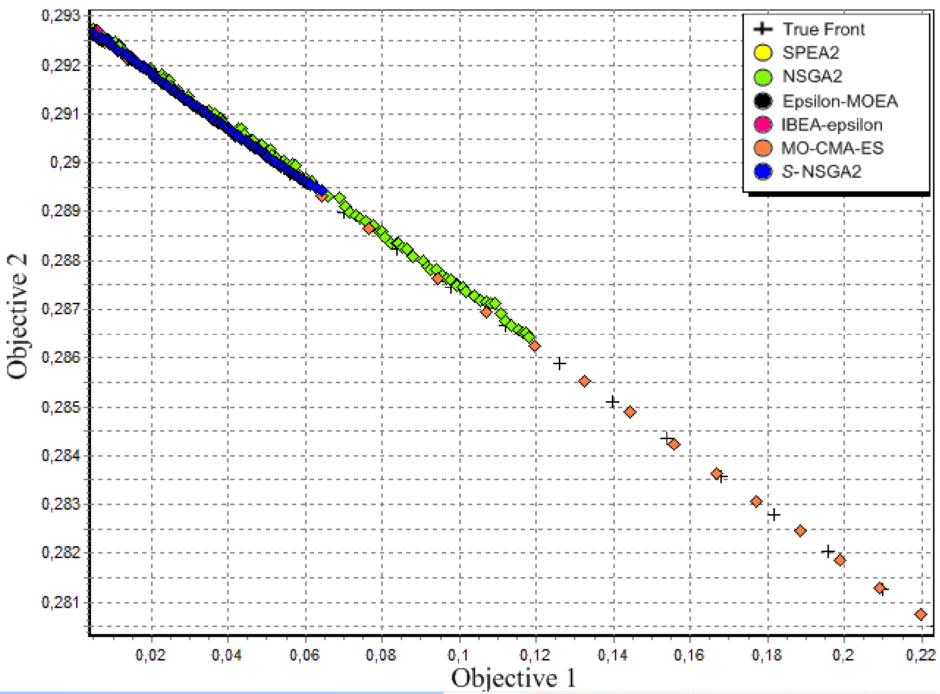
Comparison on ZDT and IHR problems

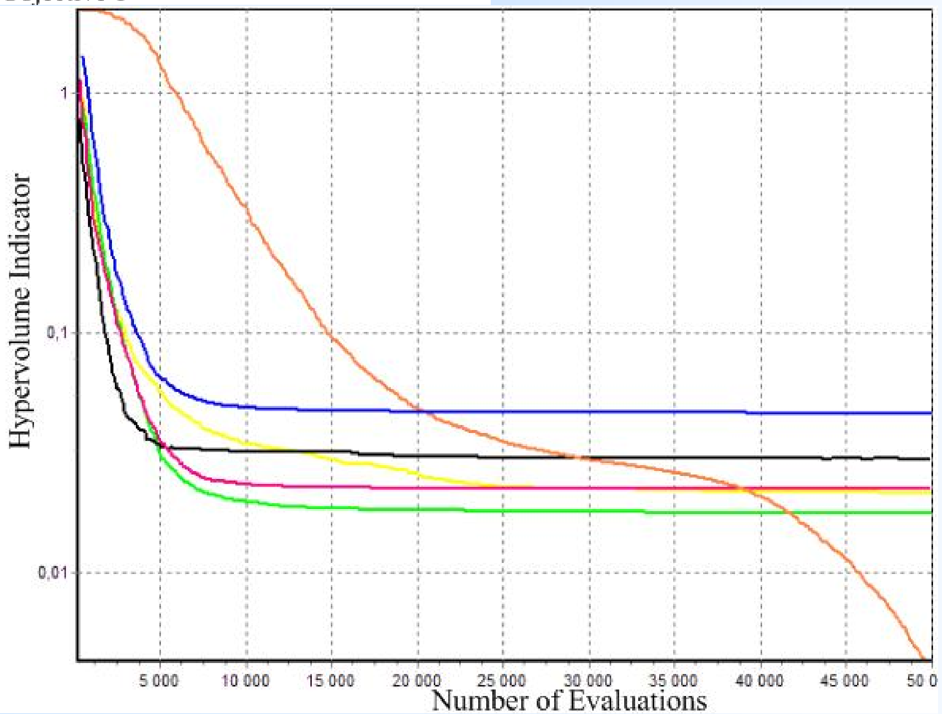
All ZDT problems :



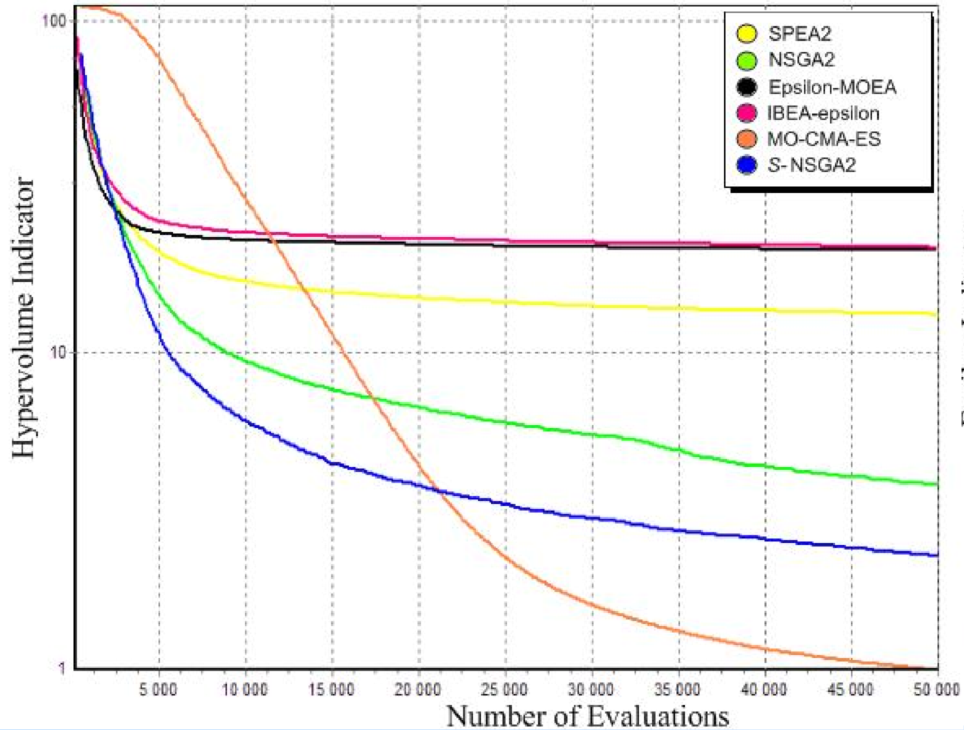


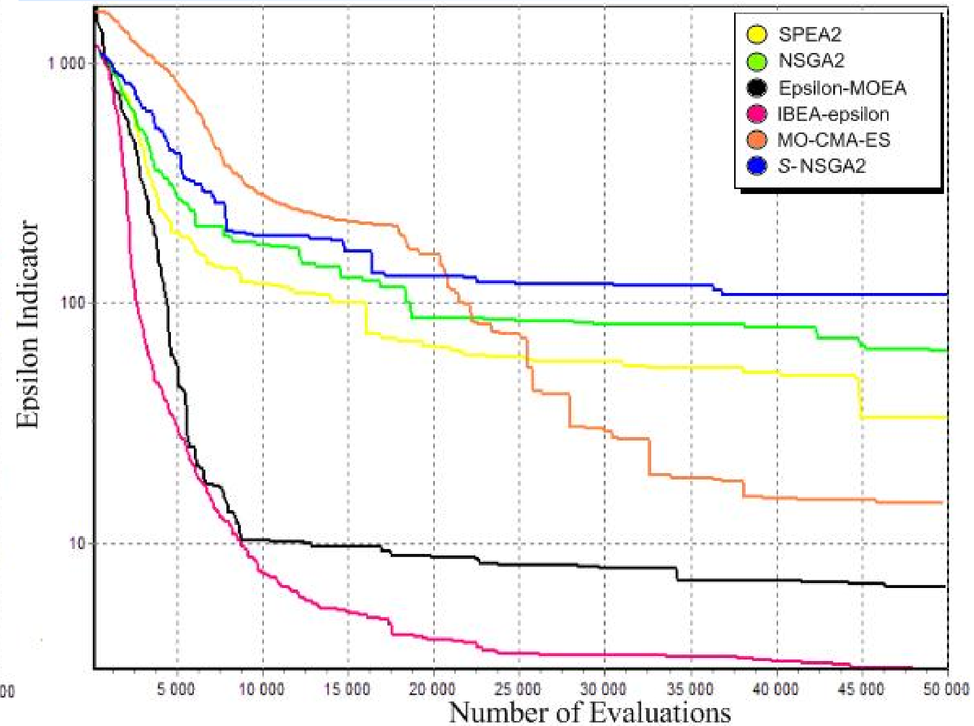
IHR1 Problem



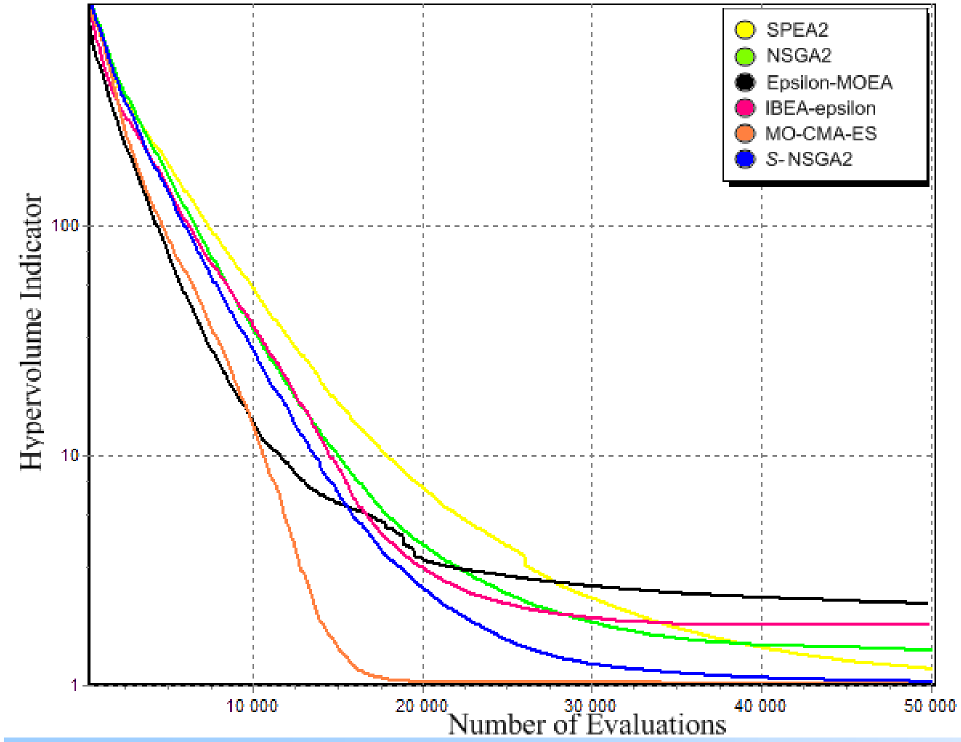


All IHR problems :

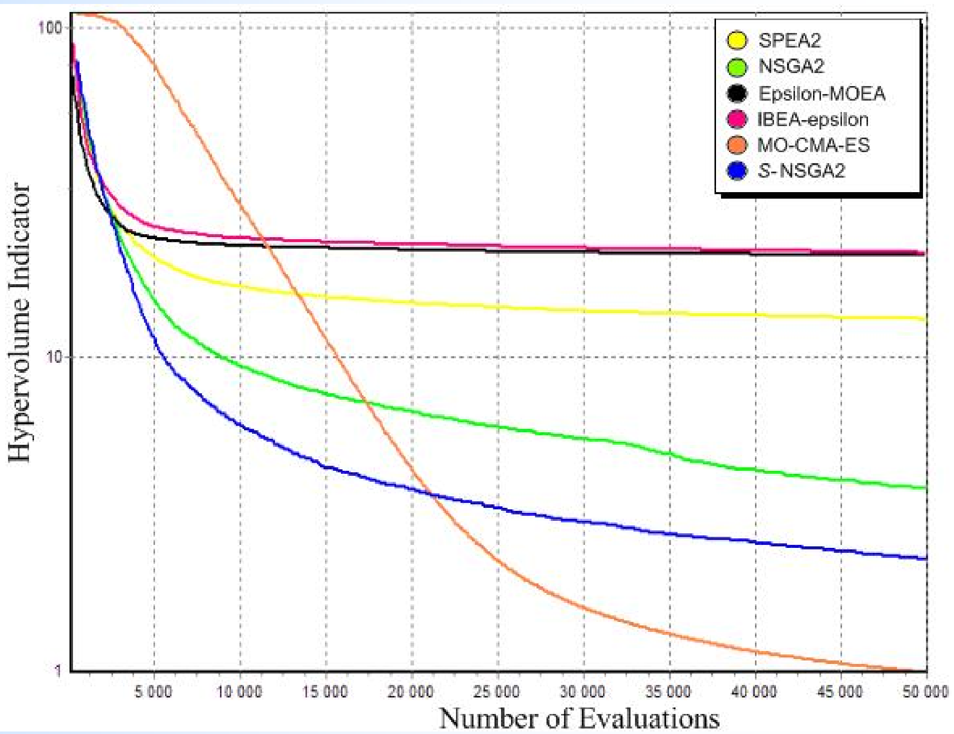




Summary :



ZDT



IHR (except IHR1)

S-NSGA2 with CMA-ES as local search procedure.

Since CMA-ES is one the most effective EA, it is reasonable to use CMA-ES as local search algorithm on the first evaluations of multiobjective optimizer. CMA-ES is the single-objective optimizer, therefore it is necessary to transform our *k*-objective problem to single-aggregate objective function :

$$g(x) = w_1 f_1(x) + \ldots + w_k f_k(x)$$

where

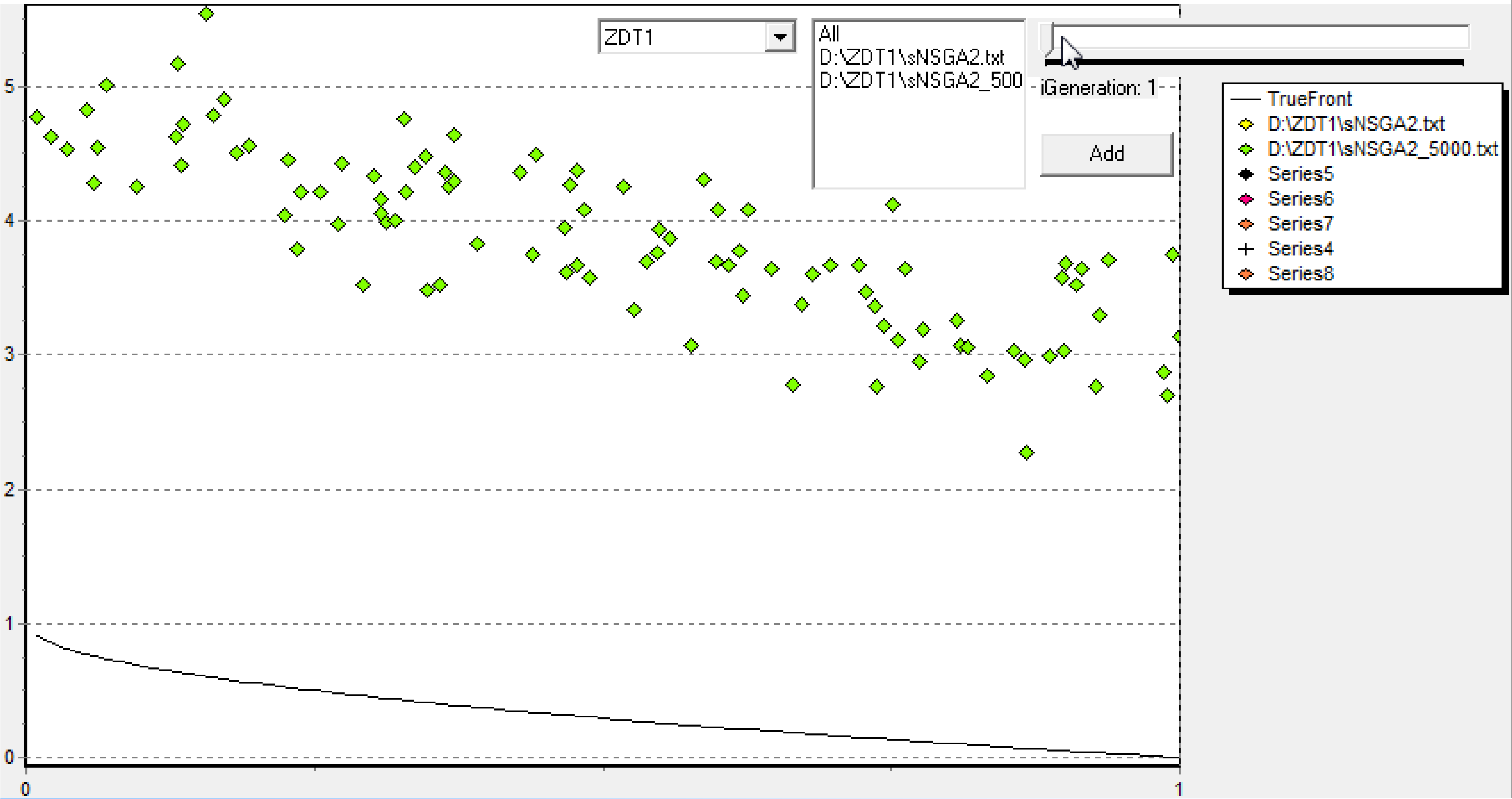
$$\sum_{i=1}^{k} w_i = 1$$

The results of CMA-ES optimization forms the first archive of multiobjective optimizer. The main problem is which stopping criterion should be used.
For all experiments on bi-objective problem we use 5000 and 7000 function evaluations as stopping criterion for CMA-ES. CMA-ES $\lambda = 30$ parents and $\mu = 10$ offspring, that is the (30,10)-CMA-ES allows to obtain good results. The optimal values of λ and μ are the topic of further study.
On bi-objective problem the aggregation simplifies to

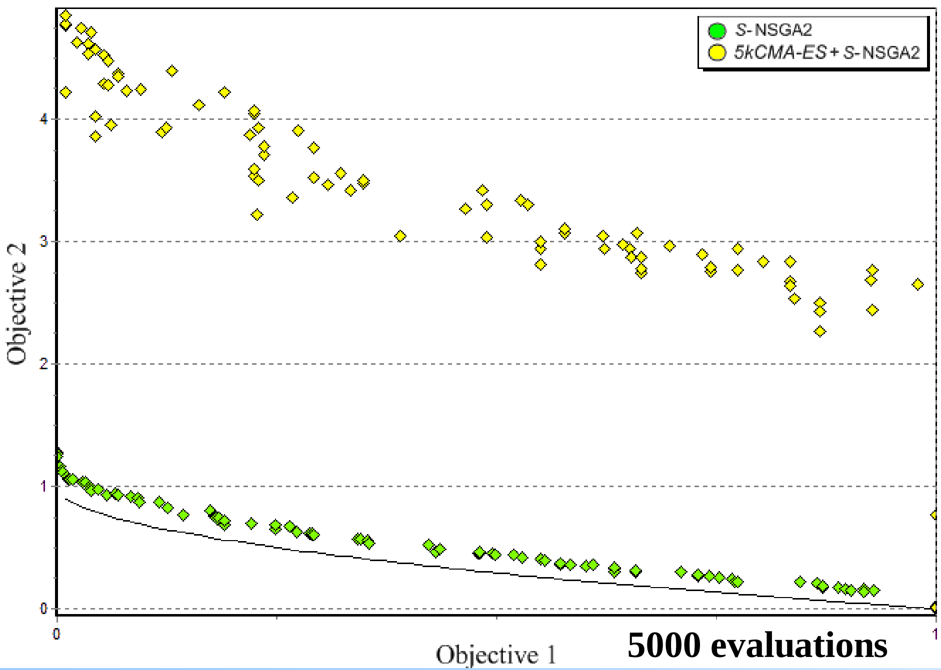
$$g(x) = \alpha f_1(x) + (1-\alpha) f_2(x)$$

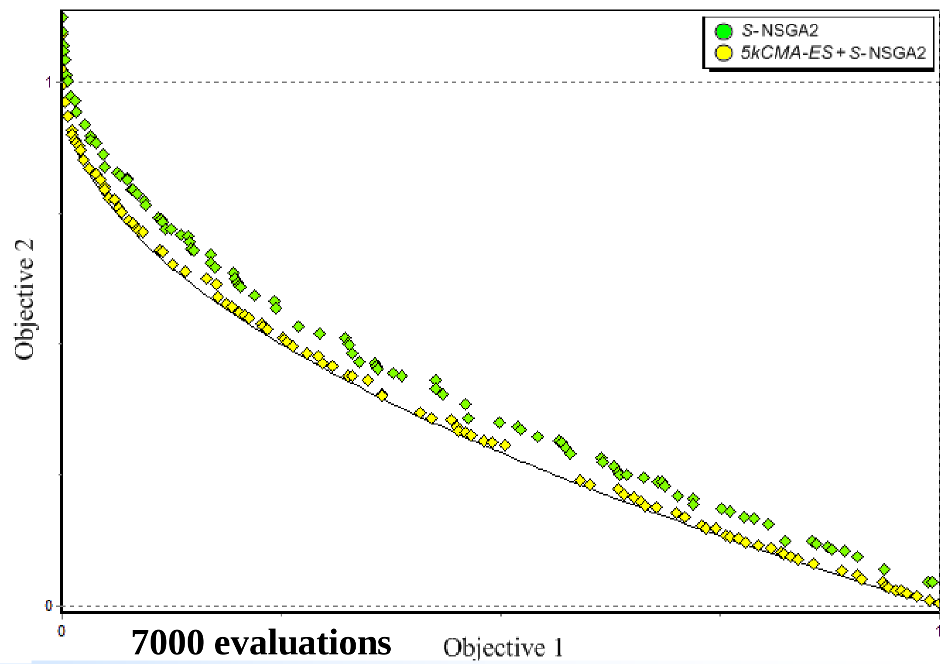
The effect of aggregation very depends on objective space and shape of Pareto-optimal front. For all experiments we use $\alpha = 0.01$.

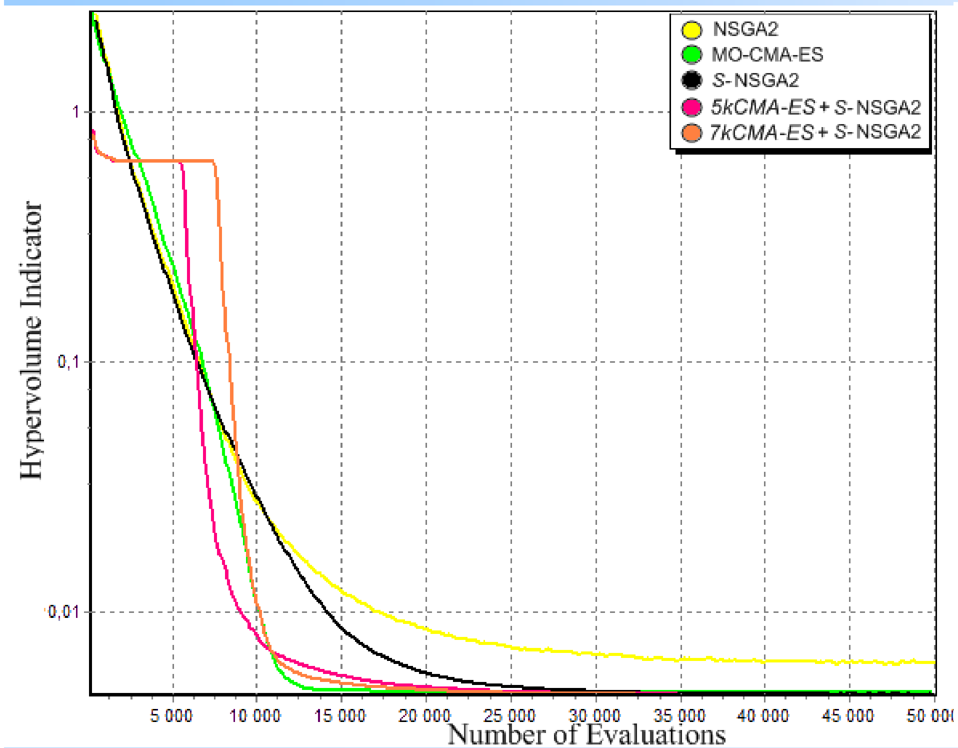
ZDT1 problem :



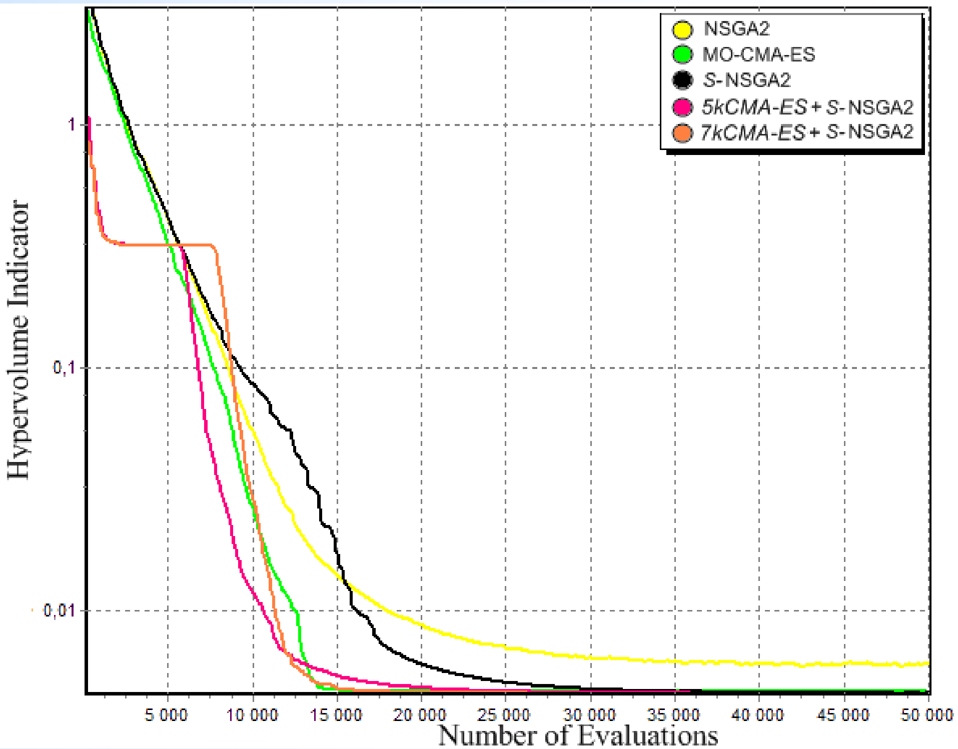
ZDT1


S-NSGA2
5kCMA-ES + S-NSGA2
Objective 2
Objective 1
0
1
2
3
4
5000 evaluations


S-NSGA2
5kCMA-ES + S-NSGA2
Objective 2
Objective 1
0
1
7000 evaluations

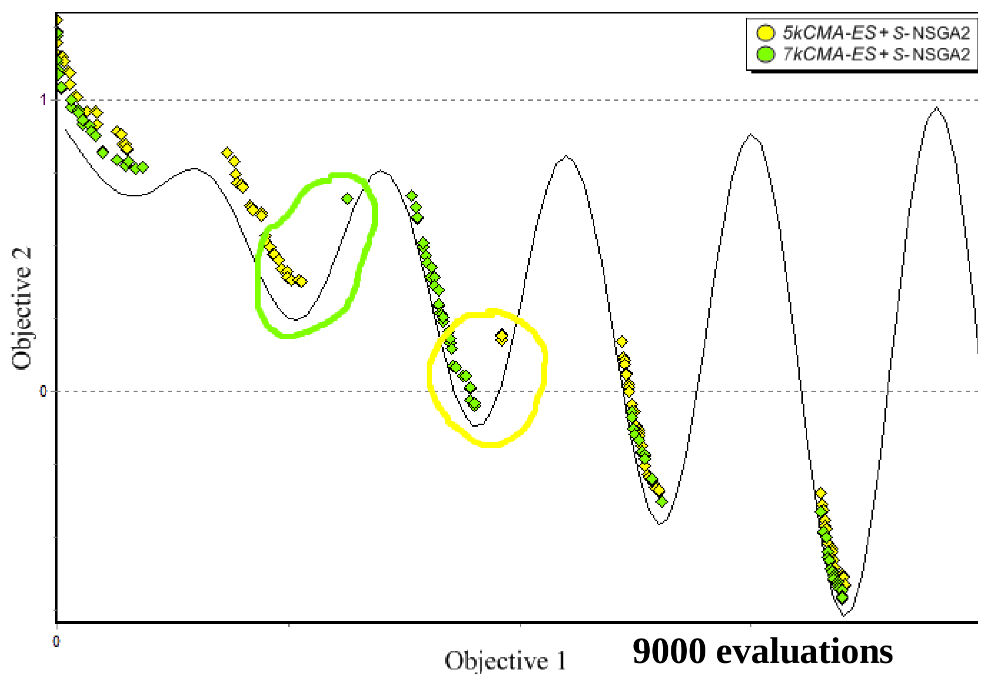

NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
Number of Evaluations

ZDT1

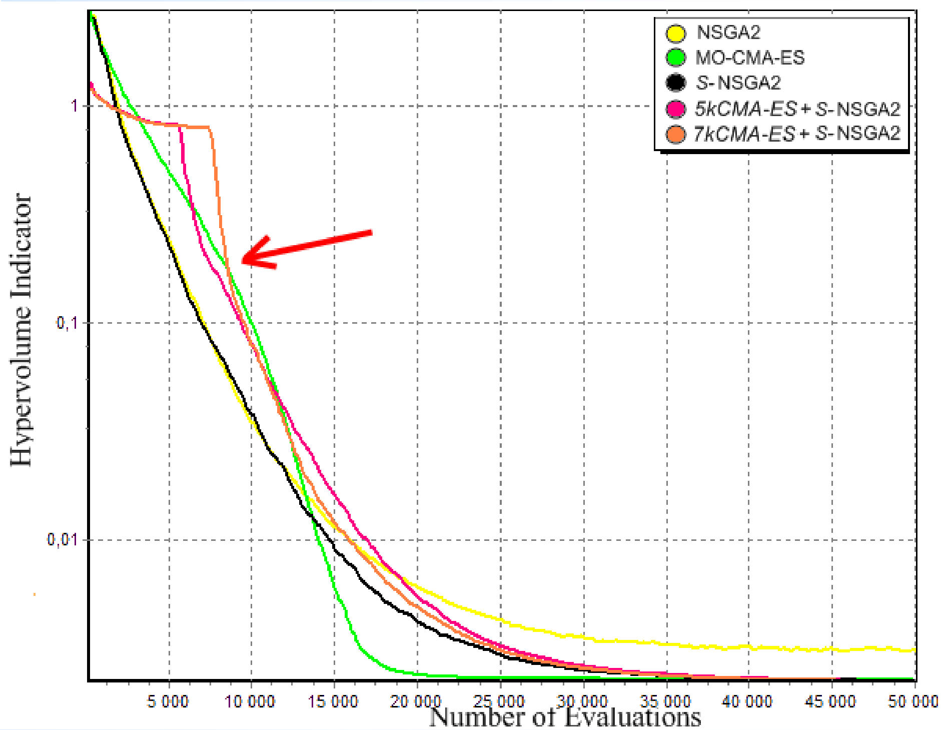

NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
Number of Evaluations

ZDT2

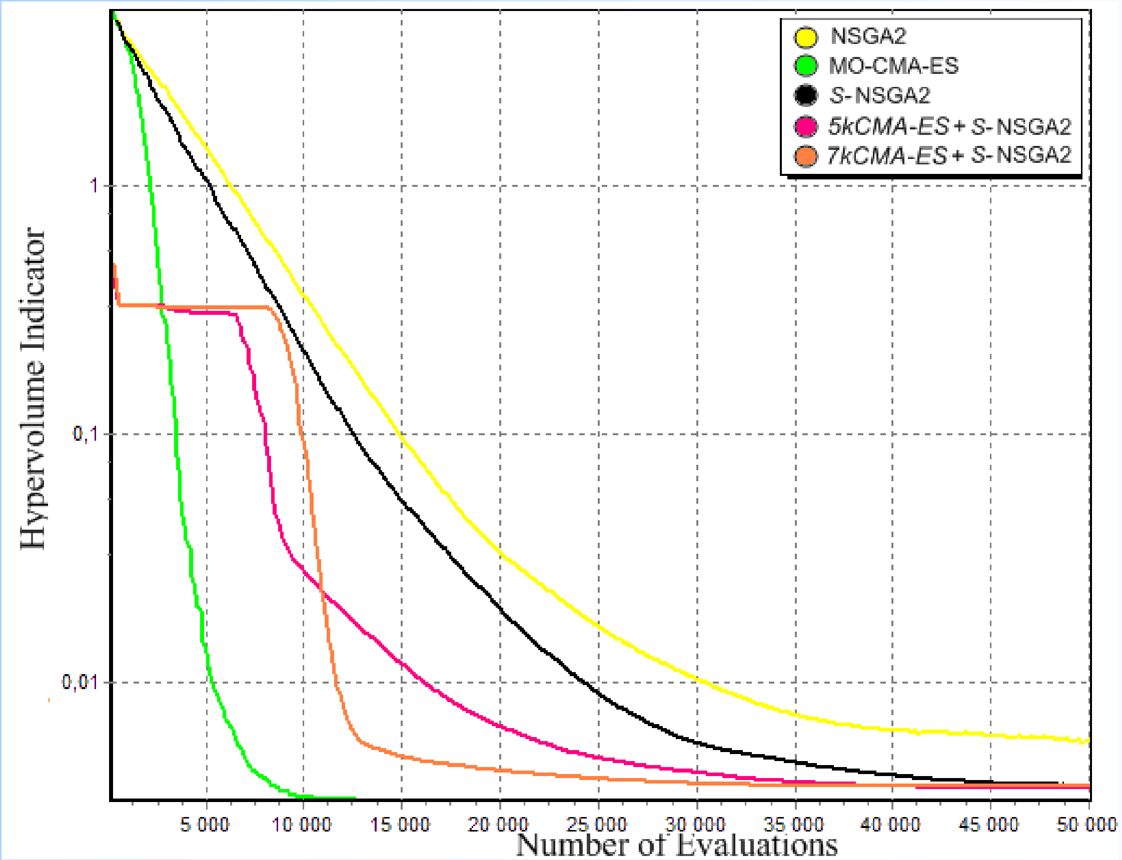
ZDT3



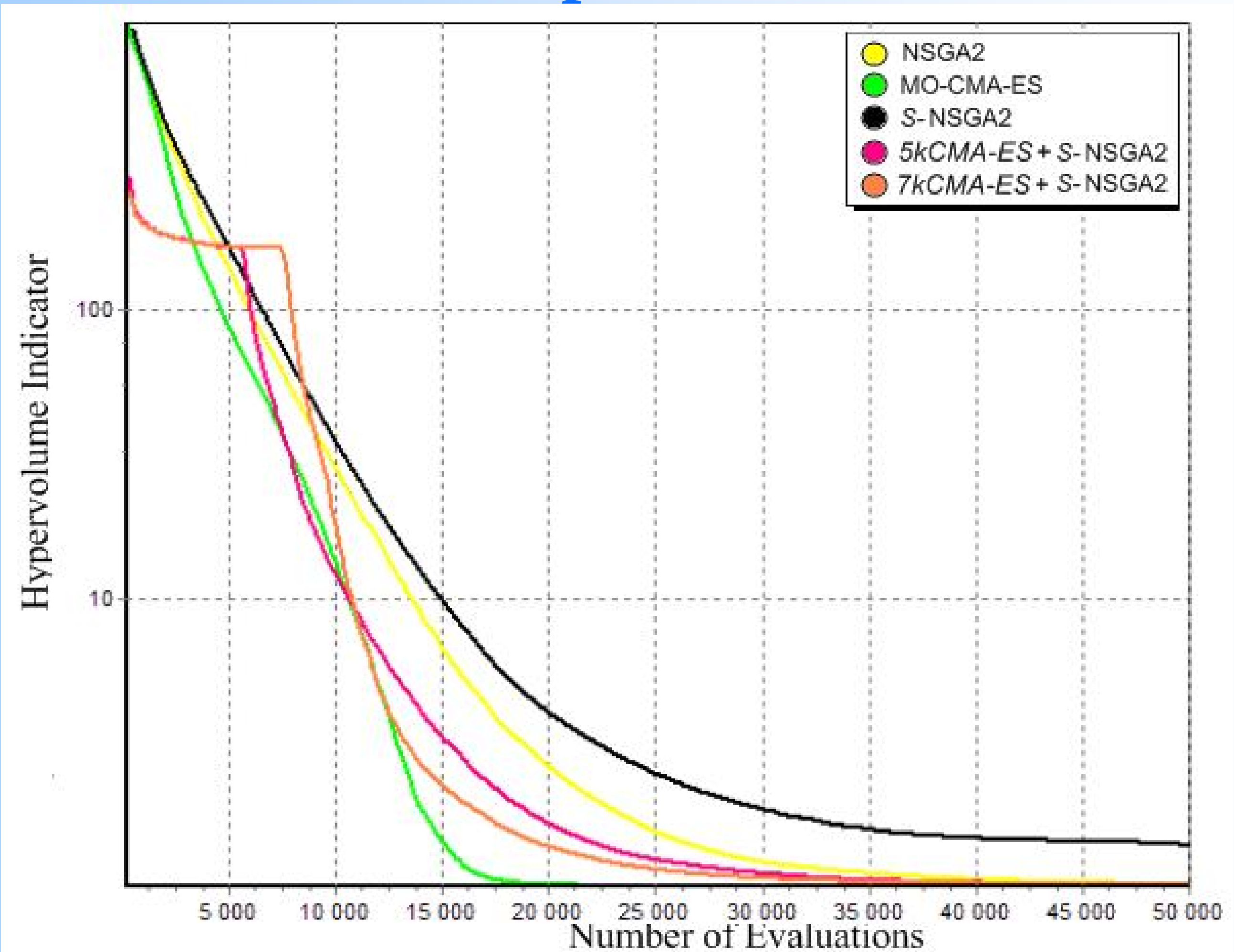
9000 evaluations

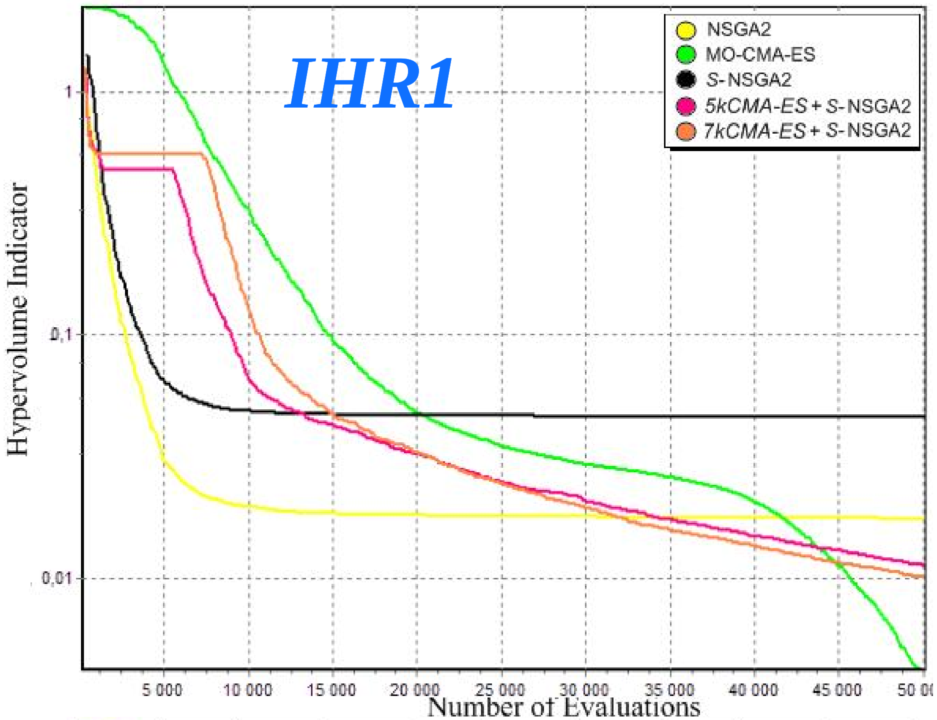


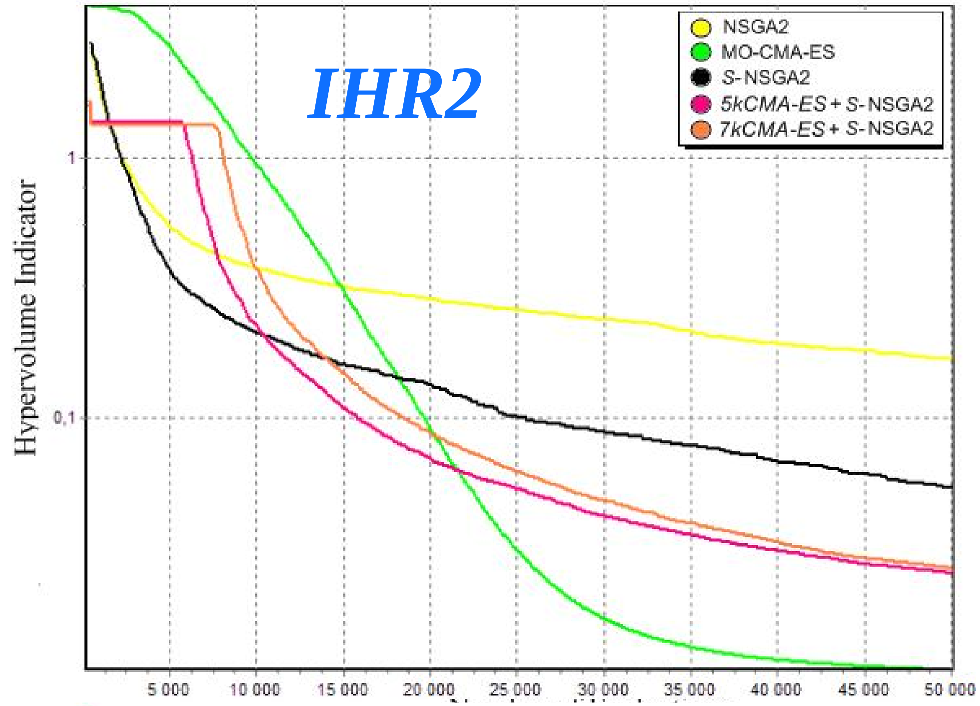
ZDT6

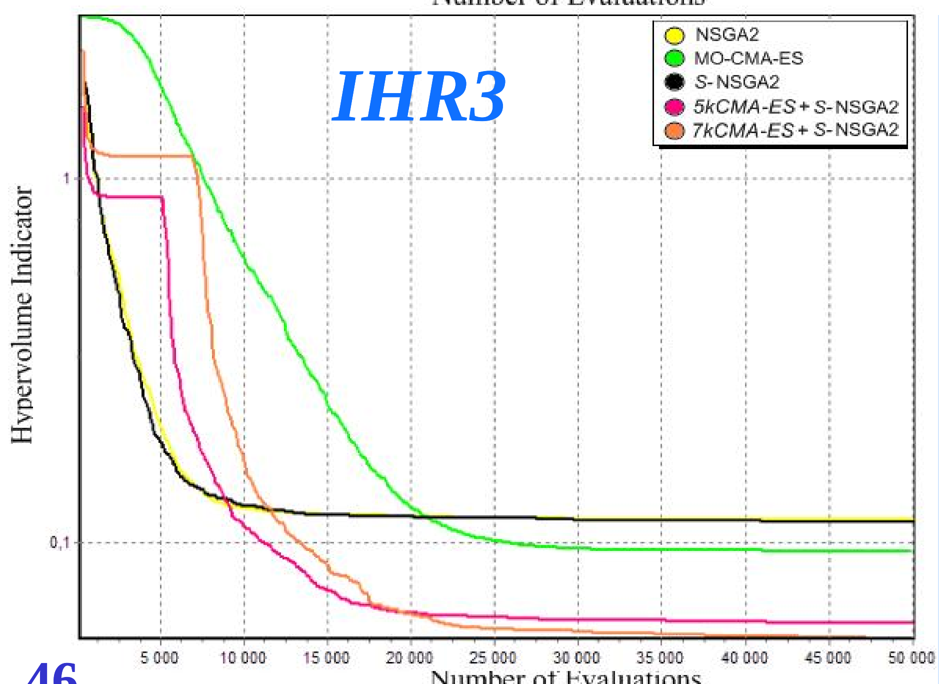

NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
1
0,1
0,01
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000
Number of Evaluations

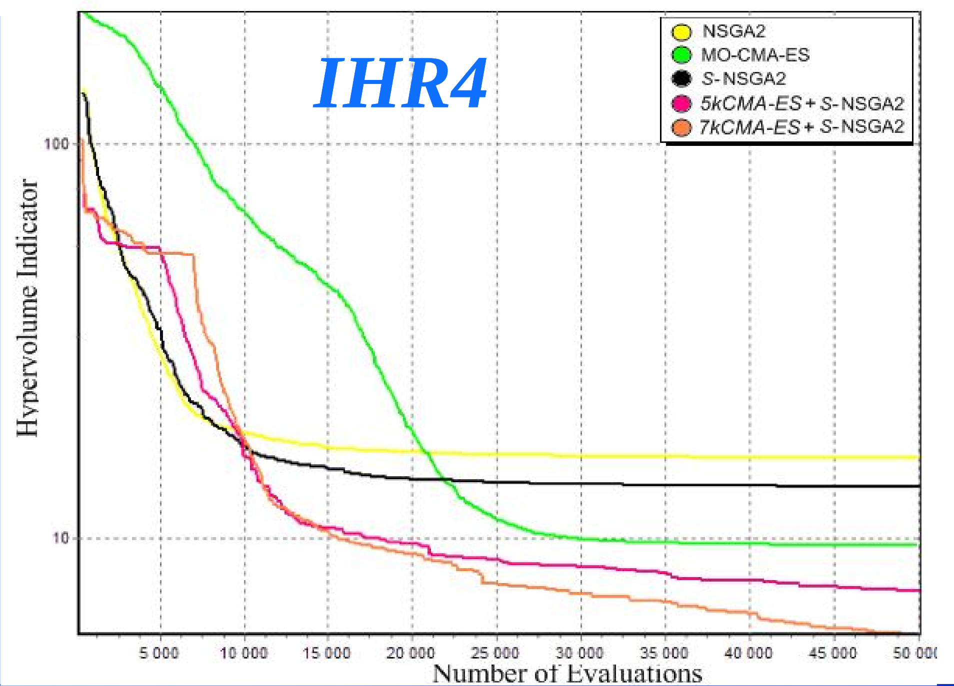
All ZDT problems



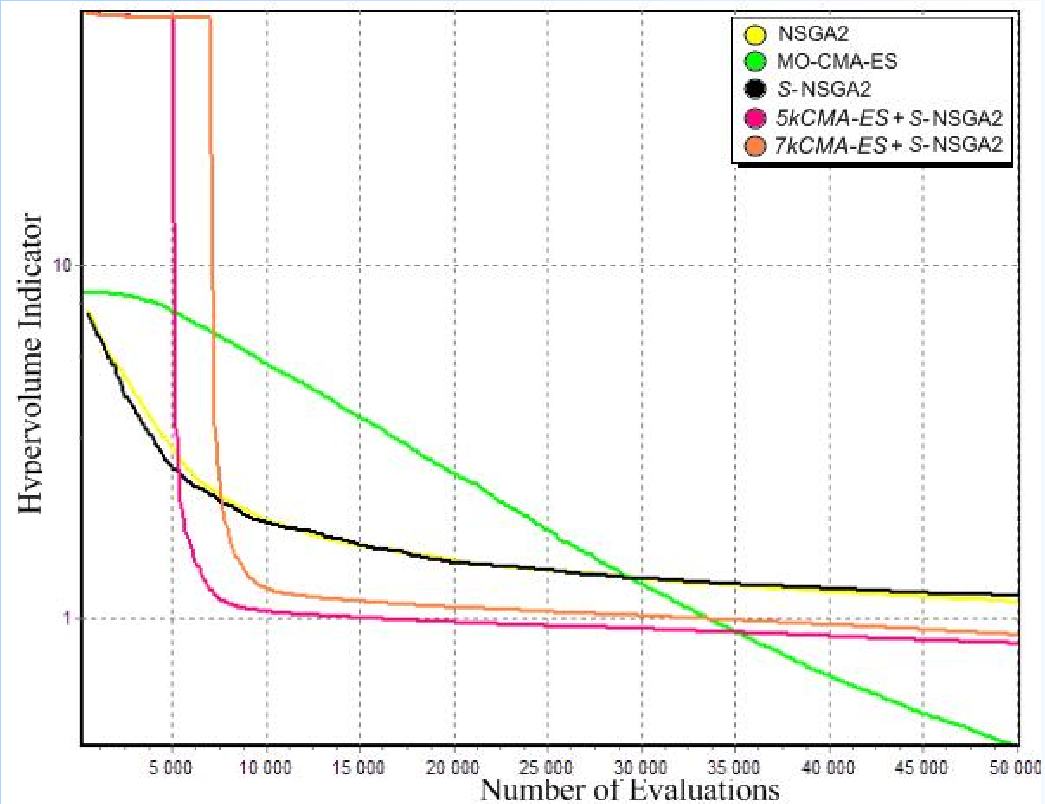

IHR1
NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
1
0,1
0,01
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
Number of Evaluations


IHR2
NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
1
0,1
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000


IHR3
NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
1
0,1
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
Number of Evaluations

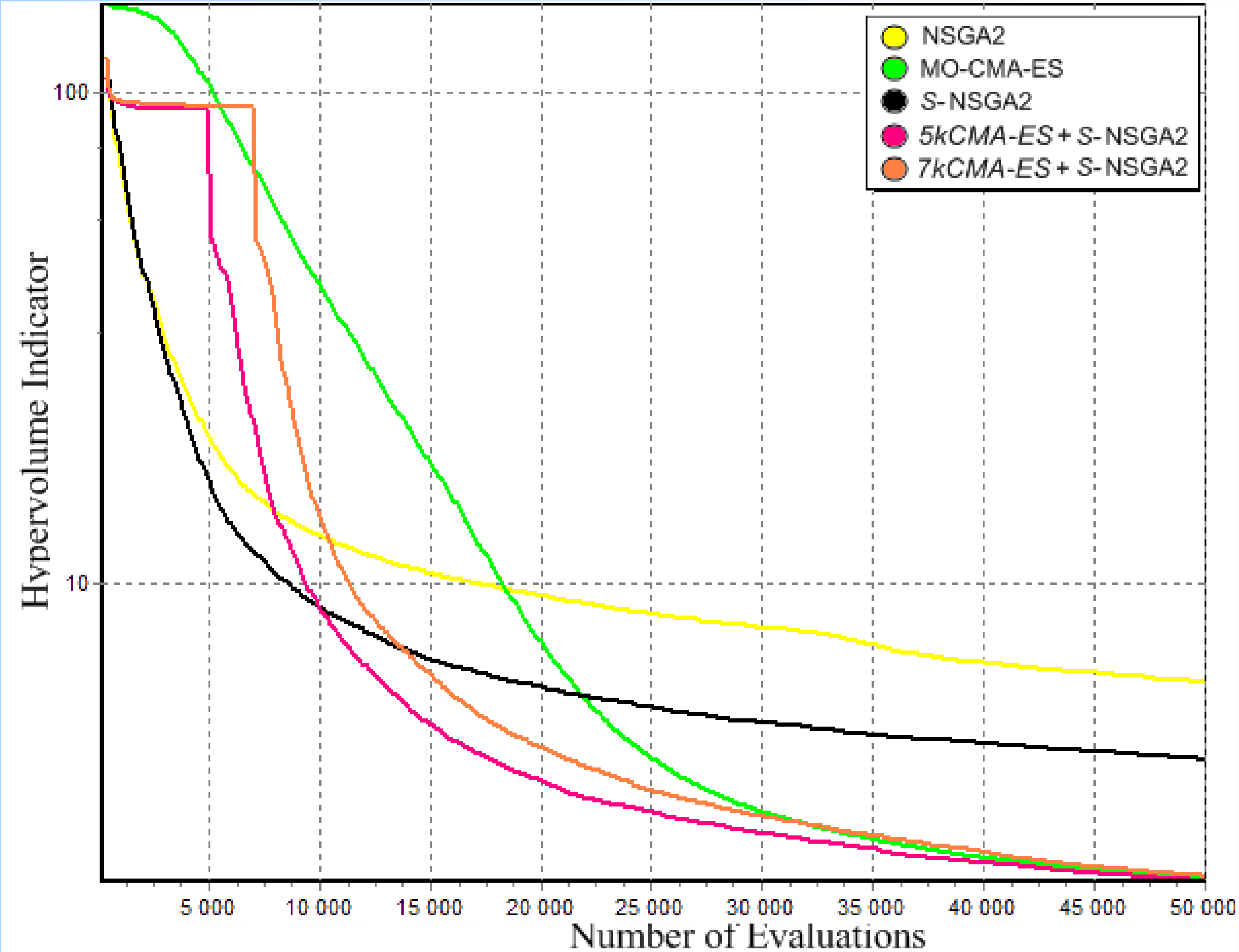

IHR4
NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
100
10
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
Number of Evaluations

IHR6

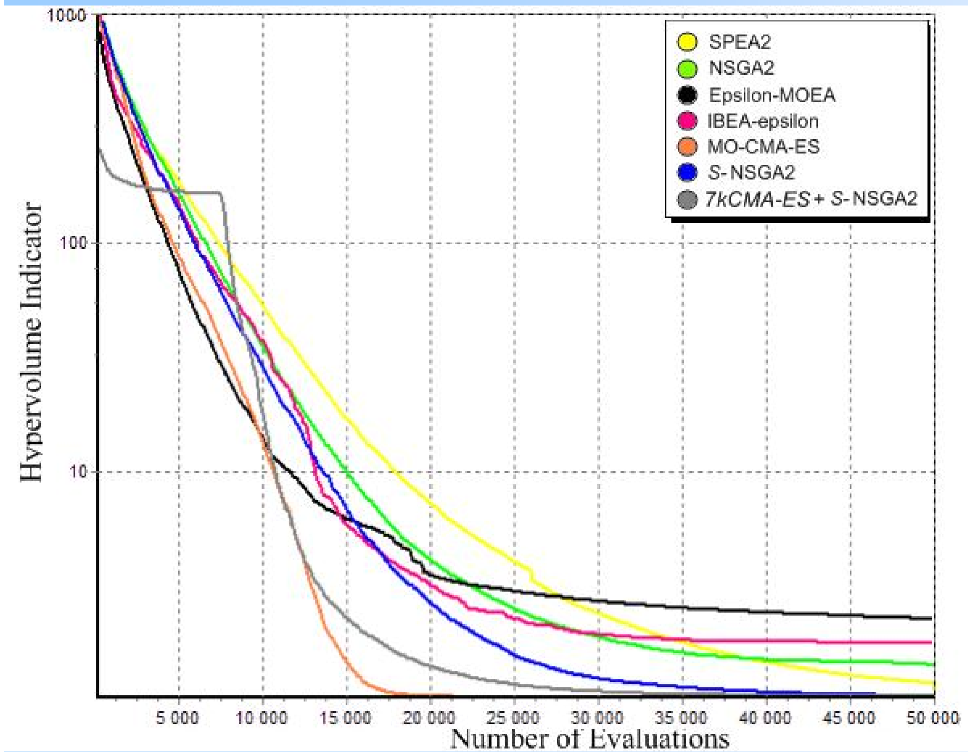


NSGA2
MO-CMA-ES
S-NSGA2
5kCMA-ES + S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
10
1
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000
Number of Evaluations

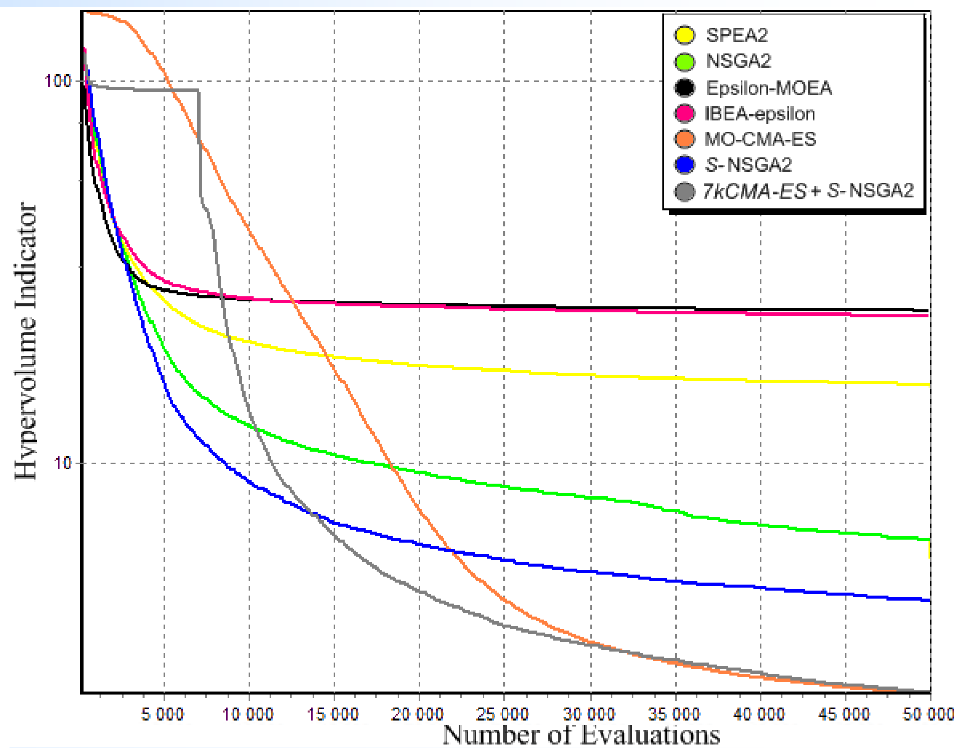
All Problems IHR



Conclusion

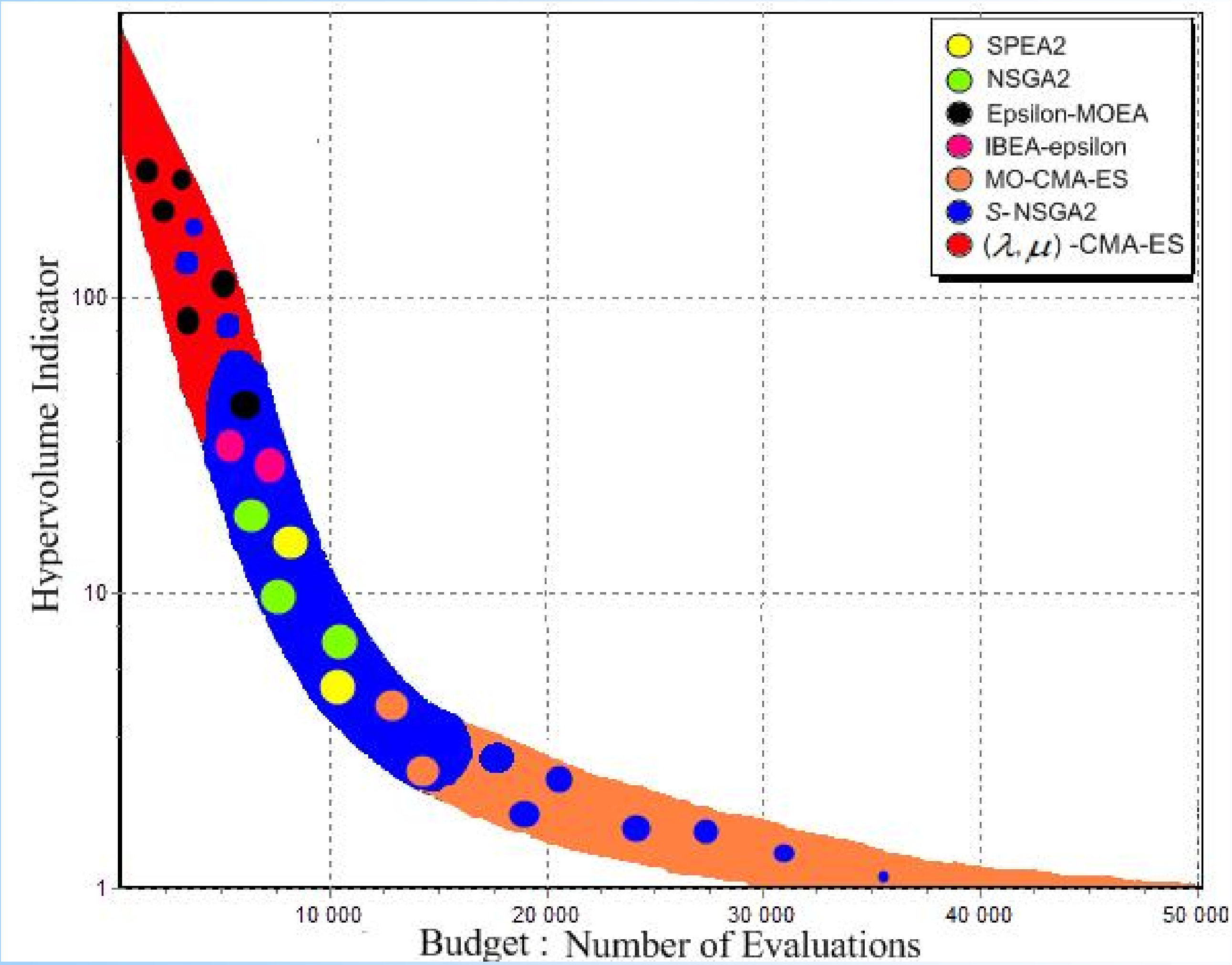


SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
1000
100
10
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000
Number of Evaluations



SPEA2
NSGA2
Epsilon-MOEA
IBEA-epsilon
MO-CMA-ES
S-NSGA2
7kCMA-ES + S-NSGA2
Hypervolume Indicator
100
10
5 000
10 000
15 000
20 000
25 000
30 000
35 000
40 000
45 000
50 000
Number of Evaluations

Some suggestion for bi-objective optimization with limited budget of function evaluations



J

Merci