

# Extension of the Pareto Active Learning Method to Multi-Objective Optimization for Stochastic Simulators

Bruno Barracosa<sup>1,2</sup>   Julien Bect<sup>2</sup>   Héloïse Dutrieux Baraffe<sup>1</sup>  
Juliette Morin<sup>1</sup>   Josselin Fournel<sup>1</sup>   Emmanuel Vazquez<sup>2</sup>

<sup>1</sup>EDF R&D, Economic and Technical Analysis of Energy Systems (EFESE), France

<sup>2</sup>Université Paris-Saclay, CNRS, CentraleSupélec, L2S, France

MASCOT-NUM Workshop on Stochastic Simulators

March 11, 2021

This work is licensed under a [Creative Commons BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.

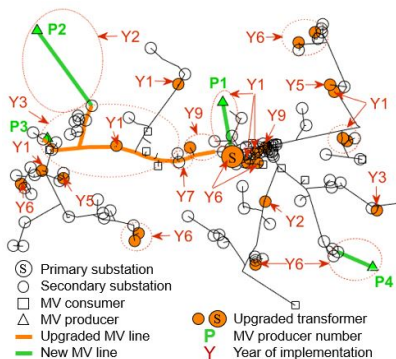


<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

- 1 Introduction
- 2 Pareto Active Learning for Stochastic Simulators
- 3 Numerical experiments
- 4 Conclusions

# Motivation

- Simulator for the multi-year electricity distribution network planning.
- Costly-to-evaluate black-box stochastic simulator (Dutrieux, 2015).



- Goal: optimize planning strategy parameters to minimize technical and economic outputs (e.g., total costs, quality of service).

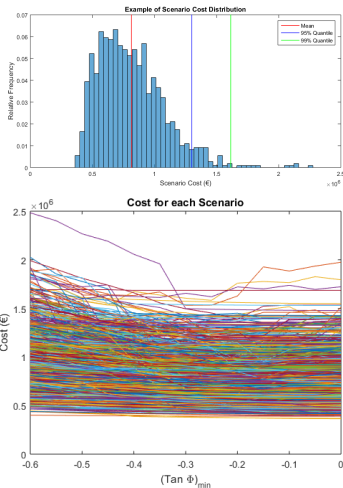
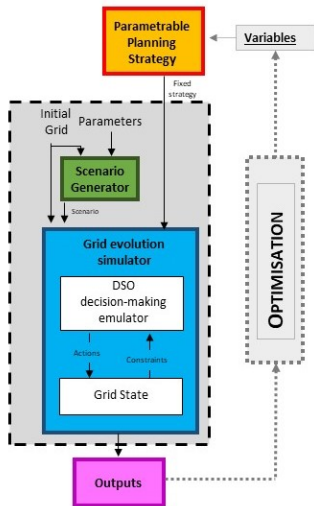
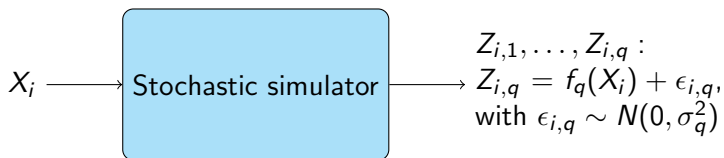


Figure: Example of cost distribution (top) and as a function of  $(\tan \Phi)_{\min}$  (bottom), per scenario.

# Problem definition

- Input: planning strategy parameters  $X_i \in \mathbb{X}$ .
- Outputs: noisy observations of latent functions  $f_1, \dots, f_q : \mathbb{X} \mapsto \mathbb{R}$ .
- Noise is additive, normally distributed and homoscedastic.



Optimization problem:

$$x^* = \arg \min_{x \in \mathbb{X}} f_1(x), \dots, f_q(x)$$

# Multi-objective optimization

Goal: identify best trade-offs among conflicting objectives.

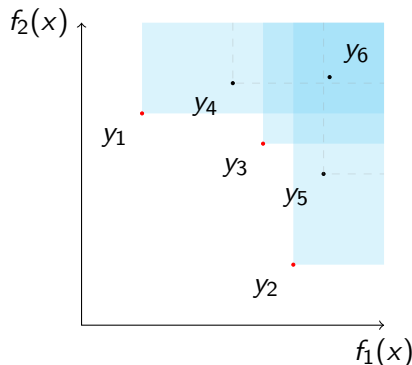
Pareto domination:  $y \prec y'$

- $y_q \leq y'_q, \forall q$
- With at least one strict inequality.

Pareto set  $\mathcal{P}$ : the set of all non-dominated points.

$$\mathcal{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, f(x') \prec f(x)\}$$

Pareto front  $\mathcal{F}$ : the image of  $\mathcal{P}$ .



**Figure:** Pareto front  $\mathcal{F} = \{y_1, y_2, y_3\}$  in a bi-objective example.

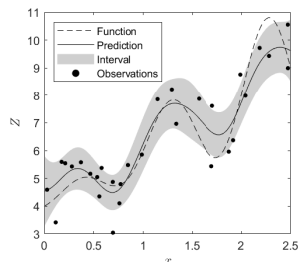
- 1 Introduction
- 2 Pareto Active Learning for Stochastic Simulators**
- 3 Numerical experiments
- 4 Conclusions



# Optimization using Gaussian processes

- Goal: select sequence of inputs to evaluate  $X_n, n = 1 \dots, N$ .
- At iteration  $n$ , previous observations  $Z_{1,q} \dots, Z_{n,q}$  used to model  $f_q$  as a sample of a GP model  $\xi_q \rightsquigarrow$  mean  $\mu_{n,q}$  and variance  $\sigma_{n,q}^2$  (prediction of  $f_q$  and uncertainty, respectively).
- GP model used to guide the optimization.

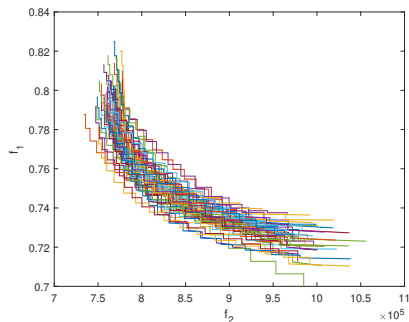
See Frazier (2018) for a tutorial.



**Figure:** GP constructed from observations (dots). Latent function (dashed), with prediction (line) and uncertainty interval (gray).

# Multi-objective noisy Bayesian optimisation

- Noise introduces observation uncertainty.
- Increased challenge in Pareto-domination assessment.

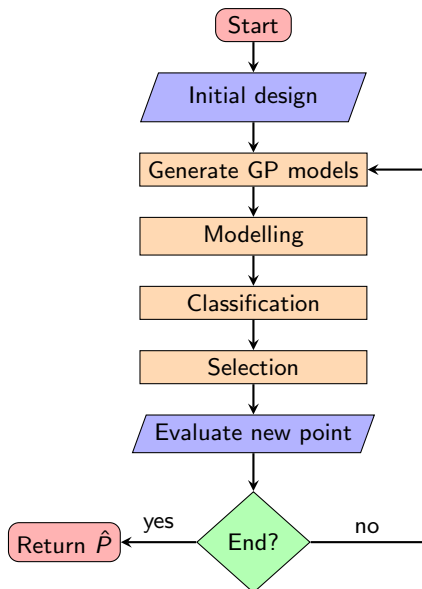


**Figure:** Example of Pareto fronts generated from noisy observations.

# PALS: Pareto Active Learning for Stochastic Simulators

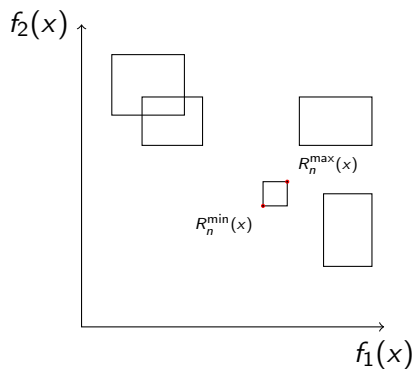
- Modification of the PAL algorithm (Zuluaga et al., 2013) to stochastic simulators.
- Strategy: classify each  $x \in \mathbb{X}$  based on a region  $R_n(x) \in \mathbb{R}^q$ .

See Barracosa, et al. (2021) for details



# PALS: Pareto Active Learning for Stochastic Simulators

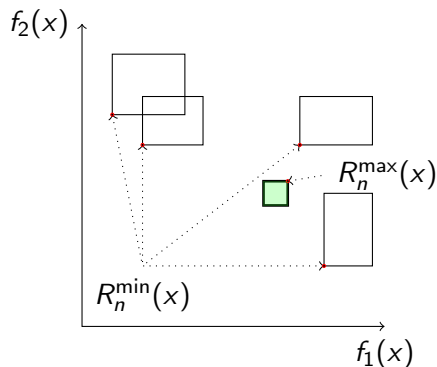
At each iteration  $n$ , each  $x$  is classified according to  $R_n(x)$ , built from GP prediction quantiles:



# PALS: Pareto Active Learning for Stochastic Simulators

At each iteration  $n$ , each  $x$  is classified according to  $R_n(x)$ , built from GP prediction quantiles:

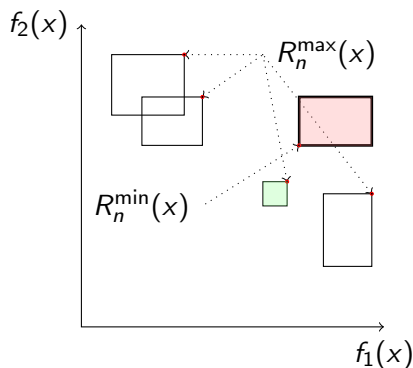
- 1  $R_n^{\max}$  of  $x$  is not dominated by another  $R_n^{\min}$ : classify  $x$  as **Pareto-optimal** ( $P_n$ ).



# PALS: Pareto Active Learning for Stochastic Simulators

At each iteration  $n$ , each  $x$  is classified according to  $R_n(x)$ , built from GP prediction quantiles:

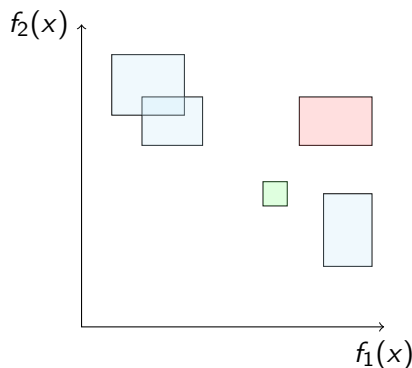
- 1  $R_n^{\max}$  of  $x$  is not dominated by another  $R_n^{\min}$ : classify  $x$  as **Pareto-optimal** ( $P_n$ ).
- 2  $R_n^{\min}$  of  $x$  is dominated by another  $R_n^{\max}$ : classify  $x$  as **non Pareto-optimal**.



# PALS: Pareto Active Learning for Stochastic Simulators

At each iteration  $n$ , each  $x$  is classified according to  $R_n(x)$ , built from GP prediction quantiles:

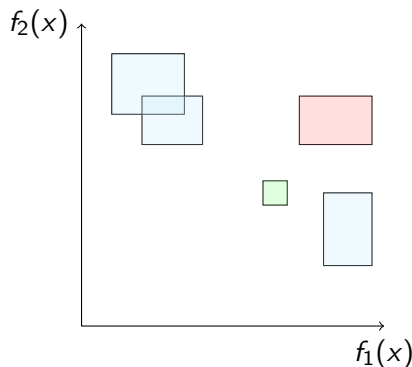
- 1  $R_n^{\max}$  of  $x$  is not dominated by another  $R_n^{\min}$ : classify  $x$  as **Pareto-optimal** ( $P_n$ ).
- 2  $R_n^{\min}$  of  $x$  is dominated by another  $R_n^{\max}$ : classify  $x$  as **non Pareto-optimal**.
- 3 Otherwise:  $x$  remains **unclassified** ( $U_n$ ).



# PALS: Pareto Active Learning for Stochastic Simulators

At each iteration  $n$ , each  $x$  is classified according to  $R_n(x)$ , built from GP prediction quantiles:

- 1  $R_n^{\max}$  of  $x$  is not dominated by another  $R_n^{\min}$ : classify  $x$  as **Pareto-optimal** ( $P_n$ ).
- 2  $R_n^{\min}$  of  $x$  is dominated by another  $R_n^{\max}$ : classify  $x$  as **non Pareto-optimal**.
- 3 Otherwise:  $x$  remains **unclassified** ( $U_n$ ).

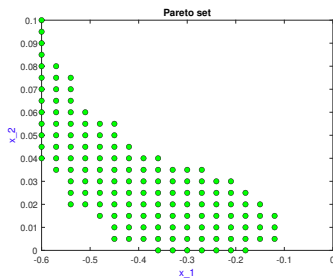
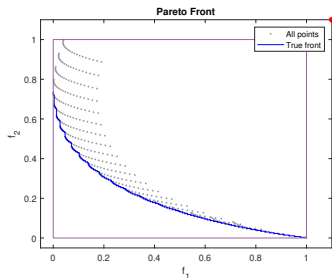
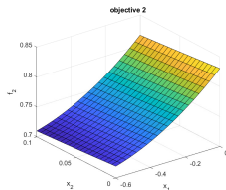
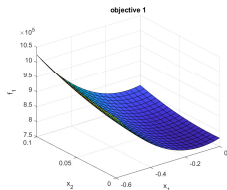


Select  $X_{n+1}$ :

$$X_{n+1} = \arg \max_{x \in (P_n \cup U_n)} \|R_n^{\min}(x) - R_n^{\max}(x)\|$$



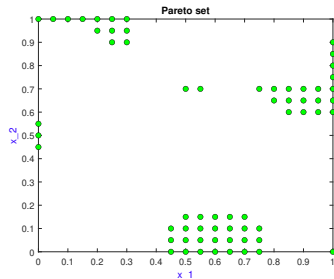
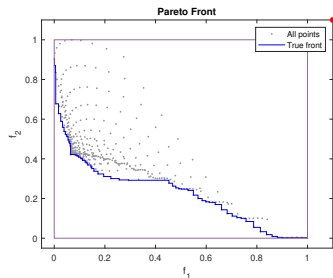
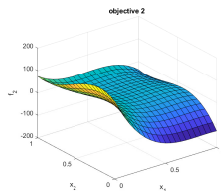
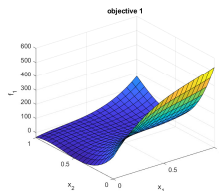
# Problem for example ( $g_1$ )



# Example

## Example 2

# Problem for example ( $g_8$ )

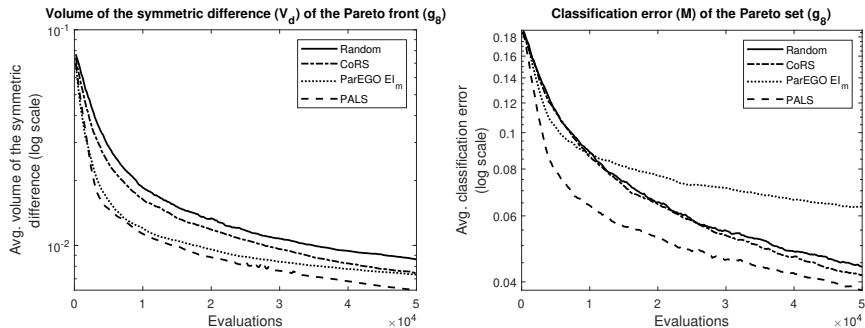


# Example 3

- 1 Introduction
- 2 Pareto Active Learning for Stochastic Simulators
- 3 Numerical experiments**
- 4 Conclusions

# Numerical experiments

- Comparison with three approaches:
  - Random.
  - “Concentrated” Random Sampling (CoRS) (Barracosa, et al., 2021).
  - ParEGO (Knowles, 2006) with  $El_m$ .
- 9 test problems:
  - Bi-objective.
  - Bi-dimensional and finite input space of size  $21 \times 21$ .
  - Homoscedastic Gaussian white noise.
- performance metrics:
  - Volume of the symmetric difference ( $V_d$ ) of the Pareto front.
  - Classification error ( $M$ ) of the Pareto set.
  - Averaged over 500 runs of the algorithm.
- Batches of 200 evaluations, and a total budget of 50,000 evaluations.



**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, a “Concentrated” Random Sampling approach, a scalarization ParEGO adapted with  $El_m$ , and PALS, for problem  $g_8$ .



**Table:** Average metrics comparison (value in percentage), at final iteration, for a Random approach, a “Concentrated” Random Sampling approach, a scalarization ParEGO adapted with  $EI_m$ , and PALS. The best metric values highlighted in bold with a green background. Metrics at 10% of the best metric are highlighted with a blue background.

$g$	Random		CoRS		$EI_m$		PALS	
	$V_d$	$M$	$V_d$	$M$	$V_d$	$M$	$V_d$	$M$
$g_1$	0.774	7.240	<b>0.630</b>	6.867	0.781	11.050	0.631	<b>5.604</b>
$g_2$	1.005	1.235	0.660	<b>0.983</b>	<b>0.628</b>	1.363	0.955	1.050
$g_3$	1.055	3.580	1.017	3.326	<b>0.710</b>	3.512	0.913	<b>3.255</b>
$g_4$	1.212	2.121	<b>1.045</b>	2.113	1.073	2.278	1.132	<b>1.934</b>
$g_5$	1.102	3.858	<b>0.662</b>	3.332	0.903	7.864	0.694	<b>3.254</b>
$g_6$	1.411	0.695	<b>0.443</b>	0.433	0.471	1.513	0.469	<b>0.387</b>
$g_7$	0.944	2.625	0.463	<b>2.531</b>	0.511	4.677	<b>0.398</b>	2.557
$g_8$	0.862	4.392	0.745	4.182	0.732	6.332	<b>0.620</b>	<b>3.809</b>
$g_9$	1.075	1.393	0.680	1.106	0.633	2.614	<b>0.562</b>	<b>0.957</b>

- 1 Introduction
- 2 Pareto Active Learning for Stochastic Simulators
- 3 Numerical experiments
- 4 Conclusions**

- PALS shows interesting performances for the multi-objective optimization of Stochastic Simulators:
  - Better performance than random approach.
  - Always good performance for Pareto set estimation.
- Future work includes:
  - Compare performance with other more complex algorithms (Picheny, 2015; Hernández-Lobato, et al. 2016).
  - Study performance impact when facing non-Gaussian and/or heteroscedastic simulators.
  - Assess performance when dealing with increased input space size or number objectives.

# References



Barracosa, B., Bect, J., Dutrieux Baraffe, H., Morin, J., Fournel, J., & Vazquez, E. (2021)

Bayesian multi-objective optimization for stochastic simulators  
*to appear soon on HAL + arXiv, available upon request*



Dutrieux, H. (2015)

Méthodes pour la planification pluriannuelle des réseaux de distribution.  
Application à l'analyse technico-économique des solutions d'intégration des énergies renouvelables intermittentes  
*Doctoral Thesis, Ecole Centrale de Lille*



Frazier, P. I. (2018)

A tutorial on Bayesian optimization  
*arXiv preprint, arXiv:1807.02811*



Hernández-Lobato, D., Hernandez-Lobato, J., Shah, A., & Adams, R. (2016)

Predictive entropy search for multi-objective bayesian optimization  
*International Conference on Machine Learning, 1492-1501*



Knowles, J. (2006)

ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems

*IEEE Transactions on Evolutionary Computation* 10(1), 50-66



Picheny, V. (2015)

Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction

*Stat Comput* 25, 1265–1280



Zuluaga, M., Krause, A., Sergent, G., & Püschel, M. (2013)

Active learning for multi-objective optimization

*30th International Conference on Machine Learning*, 462-470

# The End

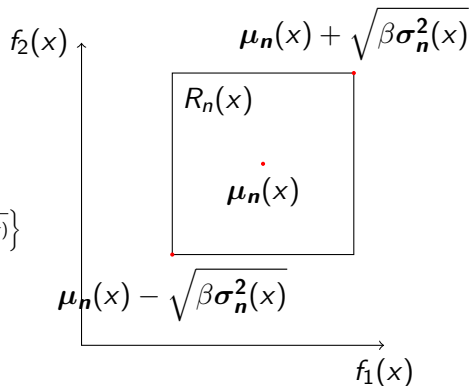
Thank you for your attention!

# Modeling<sup>1</sup>

At iteration  $n$ , GP models  $\xi_n$  used to generate predictions  $\mu_n(x)$  and prediction uncertainty  $\sigma_n^2(x)$ .

Global uncertainty represented by a region  $R_n(x)$ :

$$R_n(x) = \left\{ y \in \mathbb{R}^q : \mu_n(x) - \sqrt{\beta\sigma_n^2(x)} < y < \mu_n(x) + \sqrt{\beta\sigma_n^2(x)} \right\}$$



<sup>1</sup>Vector notation is used for simplification, e.g.,  $\mu_n(x) = (\mu_{n,1}(x), \dots, \mu_{n,q}(x))$

# Modeling<sup>1</sup>

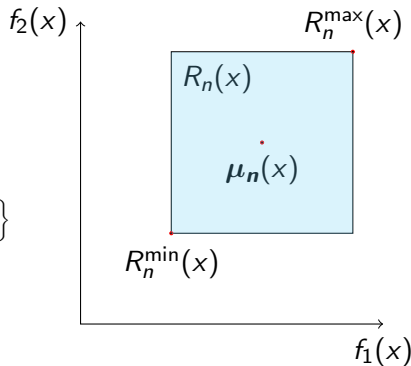
At iteration  $n$ , GP models  $\xi_n$  used to generate predictions  $\mu_n(x)$  and prediction uncertainty  $\sigma_n^2(x)$ .

Global uncertainty represented by a region  $R_n(x)$ :

$$R_n(x) = \left\{ y \in \mathbb{R}^q : \mu_n(x) - \sqrt{\beta\sigma_n^2(x)} < y < \mu_n(x) + \sqrt{\beta\sigma_n^2(x)} \right\}$$

For each  $x$  define:

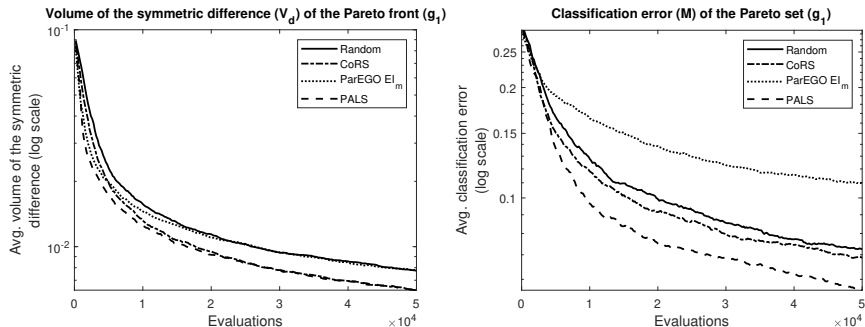
- an optimistic outcome  $R_n^{\min}(x)$ ;
- a pessimistic outcome  $R_n^{\max}(x)$ .



<sup>1</sup>Vector notation is used for simplification, e.g.,  $\mu_n(x) = (\mu_{n,1}(x), \dots, \mu_{n,q}(x))$

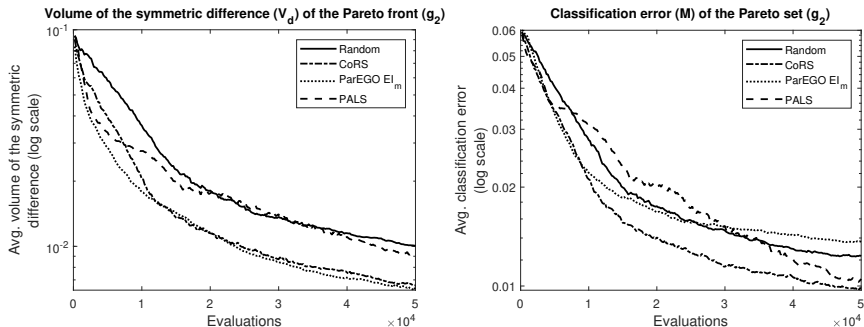


# Problem $g_1$



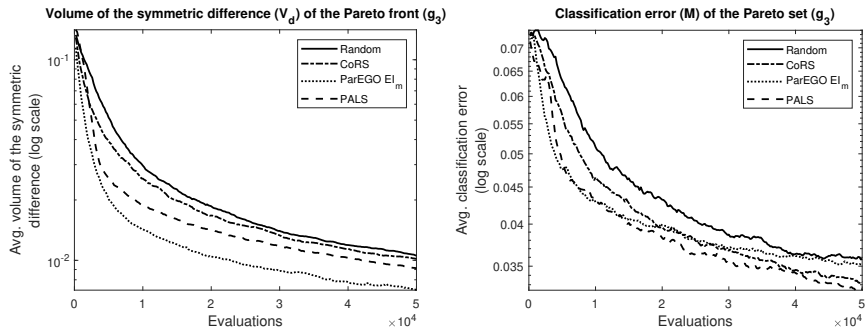
**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_1$ .

# Problem $g_2$



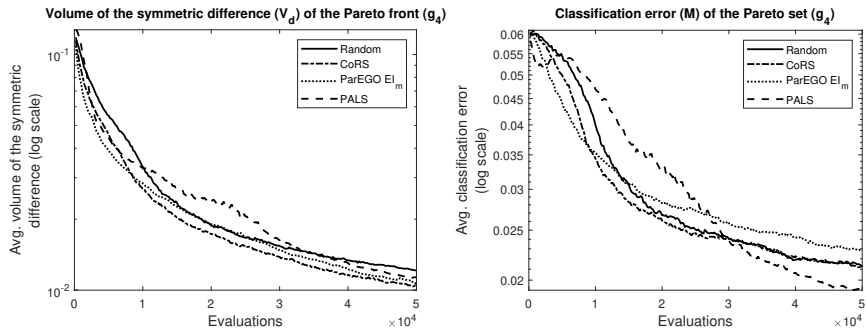
**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_2$ .

# Problem $g_3$



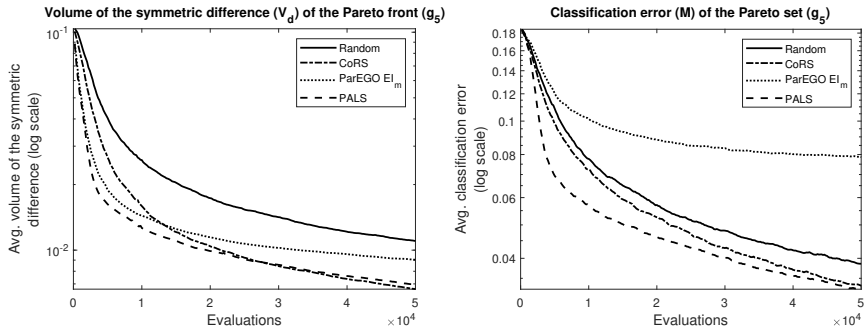
**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_3$ .

# Problem $g_4$



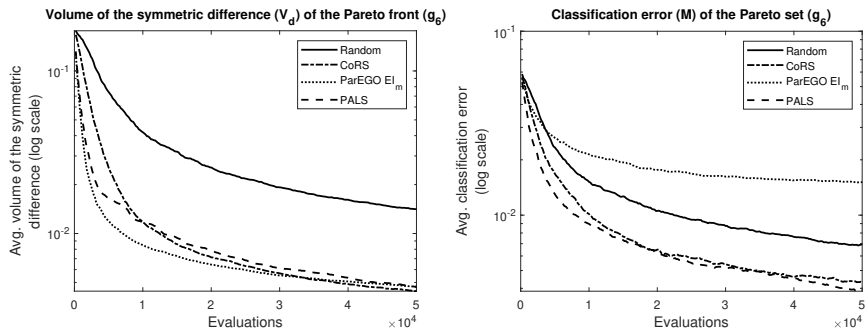
**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $EI_m$  and PALS, for problem  $g_4$ .

# Problem $g_5$



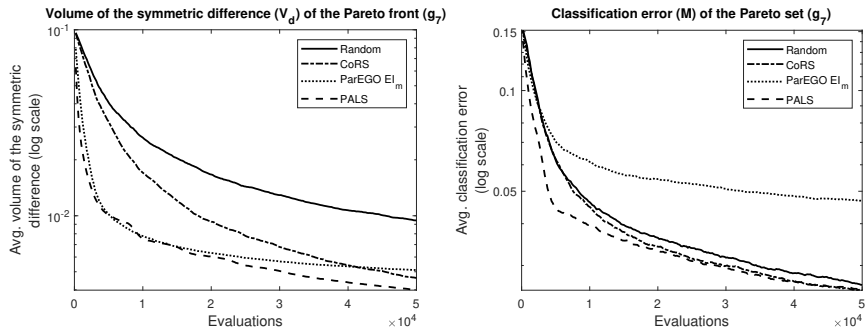
**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_5$ .

# Problem $g_6$



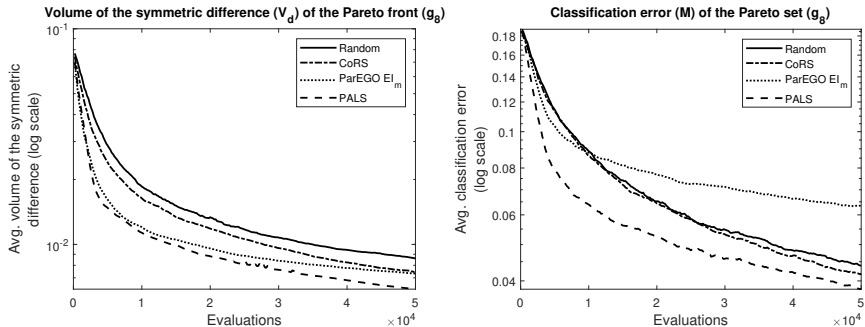
**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_6$ .

# Problem $g_7$



**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_7$ .

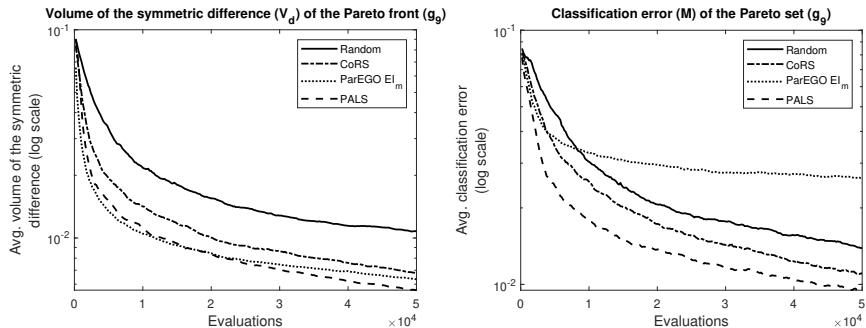
# Problem $g_8$



**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_8$ .



# Problem $g_9$



**Figure:** Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-dominance, a scalarization ParEGO adapted with  $El_m$  and PALS, for problem  $g_9$ .

**Table:** Average metrics comparison (value in percentage), at final iteration, for a Random approach, a “Concentrated” Random Sampling approach, a scalarization ParEGO adapted with  $EI_m$ , and PALS. The best metric values highlighted in bold with a green background. Metrics at 5% of the best metric are highlighted with a blue background.

$g$	Random		CoRS		$EI_m$		PALS	
	$V_d$	$M$	$V_d$	$M$	$V_d$	$M$	$V_d$	$M$
$g_1$	0.774	7.240	<b>0.630</b>	6.867	0.781	11.050	0.631	<b>5.604</b>
$g_2$	1.005	1.235	0.660	<b>0.983</b>	<b>0.628</b>	1.363	0.955	1.050
$g_3$	1.055	3.580	1.017	<b>3.326</b>	<b>0.710</b>	3.512	0.913	<b>3.255</b>
$g_4$	1.212	2.121	<b>1.045</b>	2.113	<b>1.073</b>	2.278	1.132	<b>1.934</b>
$g_5$	1.102	3.858	<b>0.662</b>	<b>3.332</b>	0.903	7.864	0.694	<b>3.254</b>
$g_6$	1.411	0.695	<b>0.443</b>	0.433	0.471	1.513	0.469	<b>0.387</b>
$g_7$	0.944	<b>2.625</b>	0.463	<b>2.531</b>	0.511	4.677	<b>0.398</b>	<b>2.557</b>
$g_8$	0.862	4.392	0.745	4.182	0.732	6.332	<b>0.620</b>	<b>3.809</b>
$g_9$	1.075	1.393	0.680	1.106	0.633	2.614	<b>0.562</b>	<b>0.957</b>

# Problem's Pareto fronts

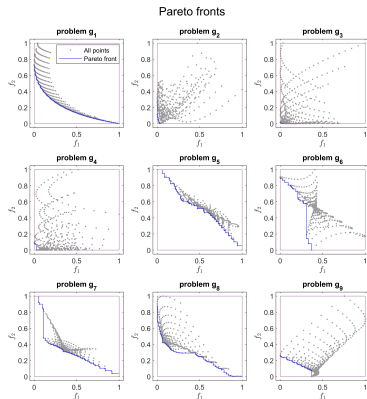


Figure: Pareto fronts of Problems  $g_1$  to  $g_9$ .

# Problem's Pareto sets

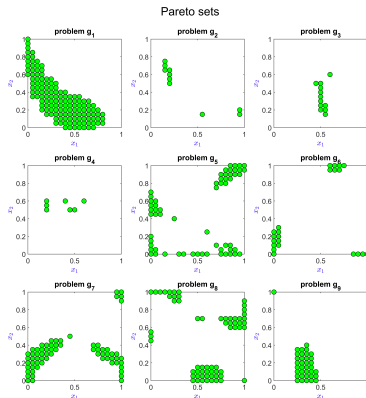


Figure: Pareto sets of Problems  $g_1$  to  $g_9$ .