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

Bruno Barracosa[1,2] Julien Bect[2] Héloise Dutrieux Baraffe[1]
Juliette Morin[1] Josselin Fournel[1] Emmanuel Vazquez[2]

[1]EDF R&D, Economic and Technical Analysis of Energy Systems (EFESE), France

[2]Université Paris-Saclay, CNRS, CentraleSupélec, L2S, France

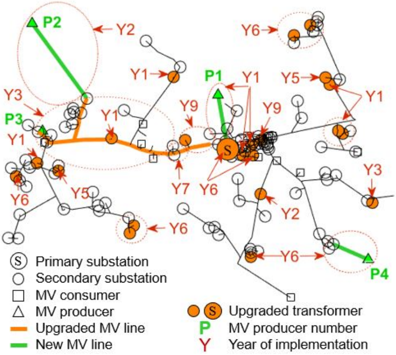
SIAM CSE 2021
MS239 Uncertainty Quantification and Optimization in Engineering

March 4, 2021

Overview

Motivation

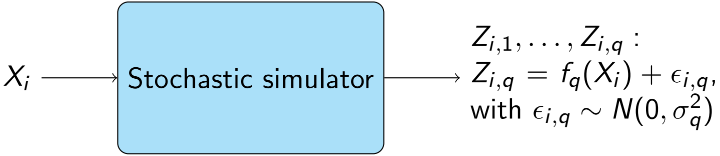
- Simulator for the multi-year electricity distribution network planning.
- Costly-to-evaluate black-box stochastic simulator (Dutrieux, 2015).



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

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:

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

Multi-objective optimization

Goal: identify best trade-offs among conflicting objectives.

Pareto domination: $y \prec y'$, when $y_q \leq y'_q, \forall q$, with at least one strict inequality.

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

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

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

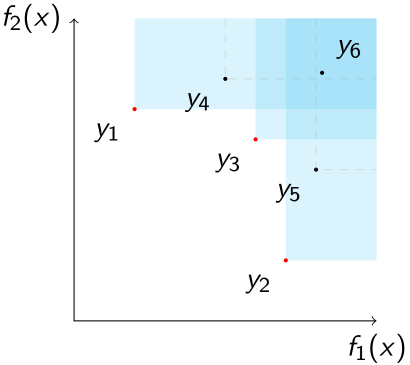


Figure: Pareto front $\mathcal{F} = \{y_1, y_2, y_3\}$ in a bi-objective example.

Overview

Optimization using Gaussian processes

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

See Frazier (2018) for a tutorial.

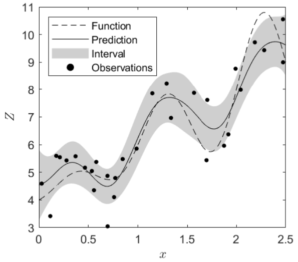
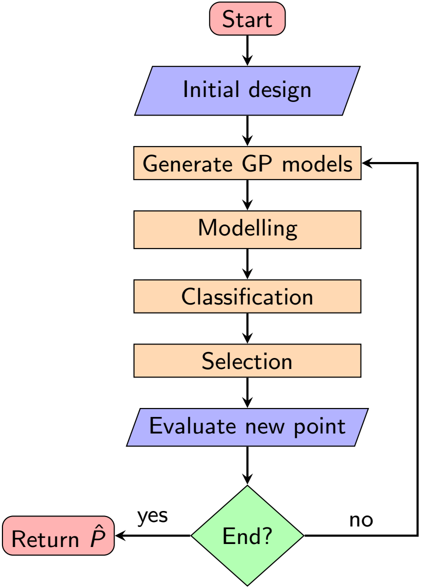


Figure: GP constructed from observations (dots). Latent function (dashed), with prediction (line) and uncertainty interval (gray).

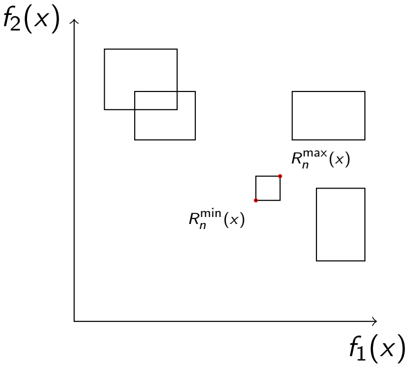
PALS: Pareto Active Learning for Stochastic Simulators

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



PALS: Pareto Active Learning for Stochastic Simulators

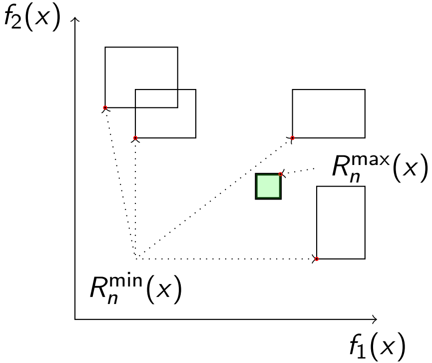
At each iteration n, each x is classified according to $R_n(x)$, built from GP prediction quantiles:



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:

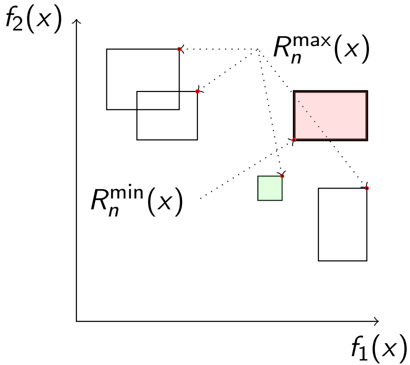
1. $R_n^{\max}$ of x is not dominated by another $R_n^{\min}$: classify x as Pareto-optimal (P_n).



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:

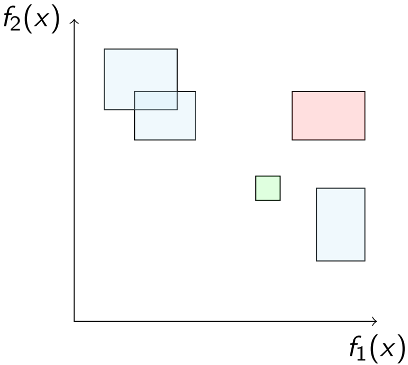
1. $R_n^{\max}$ of x is not dominated by another $R_n^{\min}$: classify x as Pareto-optimal (P_n).
2. $R_n^{\min}$ of x is dominated by another $R_n^{\max}$: classify x as non Pareto-optimal.



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:

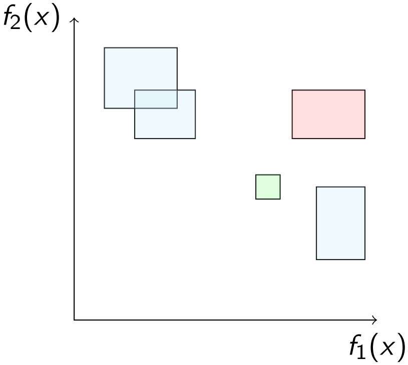
1. $R_n^{\max}$ of x is not dominated by another $R_n^{\min}$: classify x as Pareto-optimal (P_n).
2. $R_n^{\min}$ of x is dominated by another $R_n^{\max}$: classify x as non Pareto-optimal.
3. Otherwise: x remains unclassified (U_n).



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}:

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

Overview

Numerical experiments

- Comparison with three approaches:
  - Random.
  - "Concentrated" Random Sampling (CoRS).
  - ParEGO (Knowles, 2006) with EI_m.
- 9 test problems:
  - bi-objective.
  - bi-dimensional and finite input space of size 21×21.
  - 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.

Results

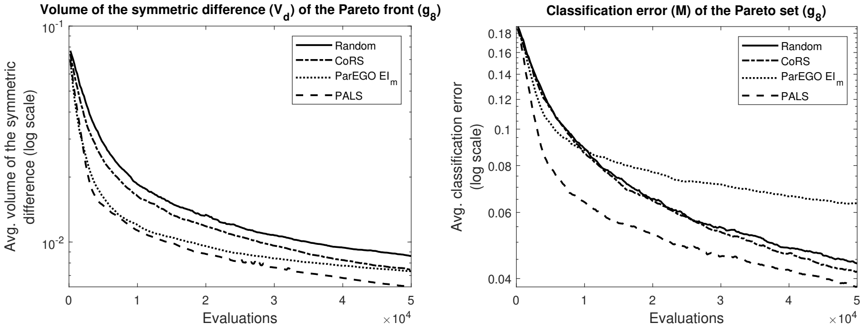


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, a “Concentrated” Random Sampling approach, a scalarization ParEGO adapted with EI_m, and PALS, for problem g_8.

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.

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

Overview

Conclusions and future work

- 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.
  - Study performance impact when facing non-Gaussian and/or heteroscedastic simulators.
  - Assess performance when dealing with increased input space size or number objectives.

References

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

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

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

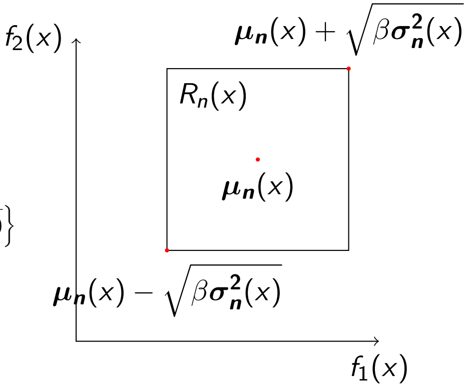
Thank you for your attention!

Modeling[1]

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

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

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



[1] Vector notation is used for simplification, e.g., $\boldsymbol{\mu_n}(x) = (\mu_{n,1}(x), \ldots, \mu_{n,q}(x))$

Modeling[1]

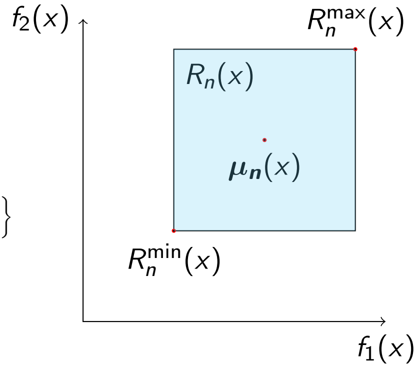
At iteration n, GP models ξ_n used to generate predictions $\boldsymbol{\mu_n}(x)$ and prediction uncertainty $\boldsymbol{\sigma_n^2}(x)$.

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

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

For each x define:

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



[1] Vector notation is used for simplification, e.g., $\boldsymbol{\mu_n}(x) = (\mu_{n,1}(x), \ldots, \mu_{n,q}(x))$

Problem g_1

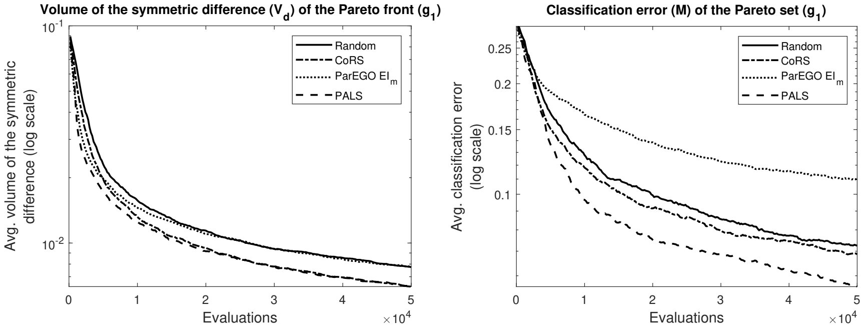


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_1.

Problem g_2

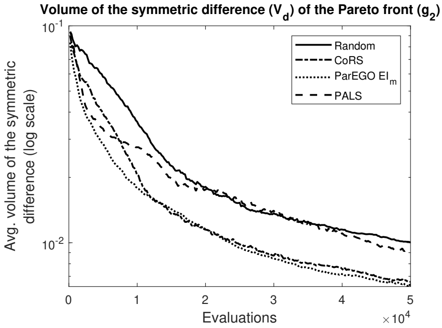


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_2.

Problem g_3

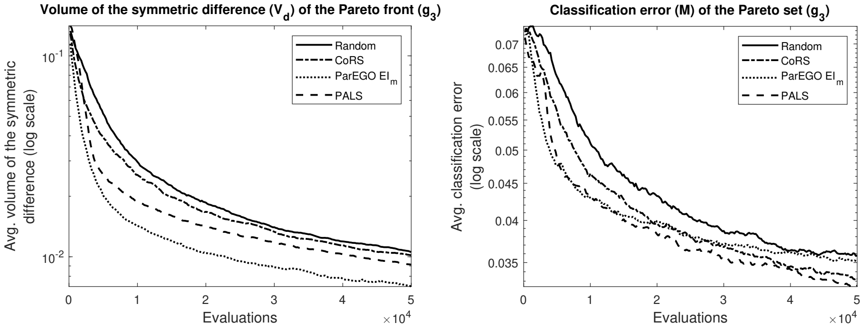
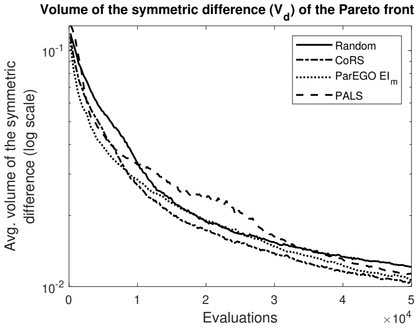


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_3.

Problem g_4



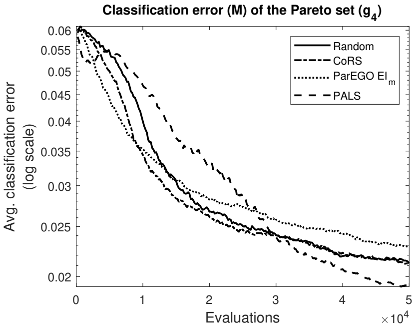


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_4.

Problem g_5

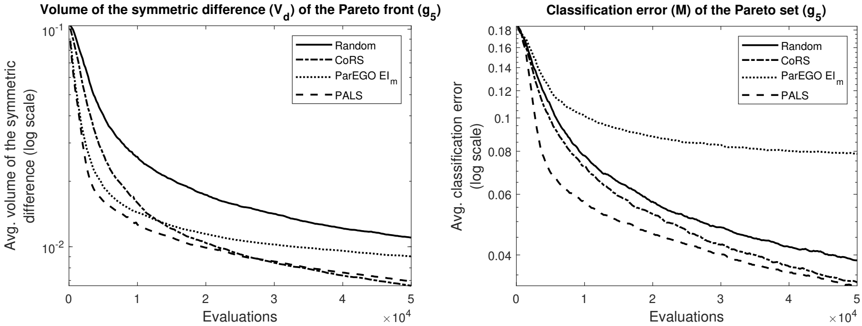


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_5.

Problem g_6

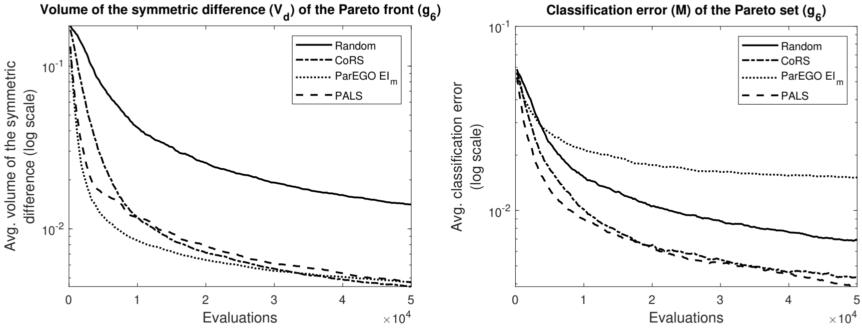


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_6.

Problem g_7

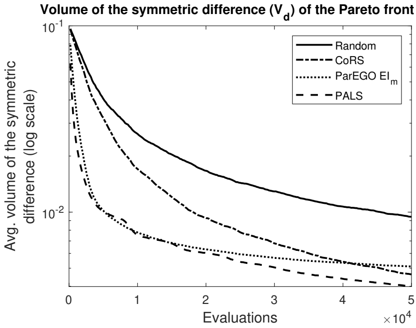


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_7.

Problem g_8

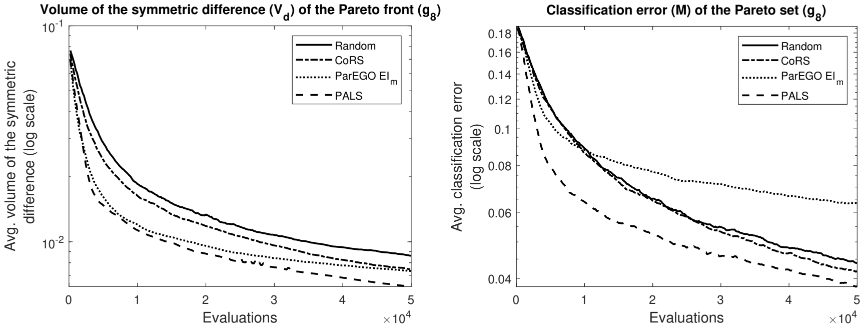


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_8.

Problem g_9

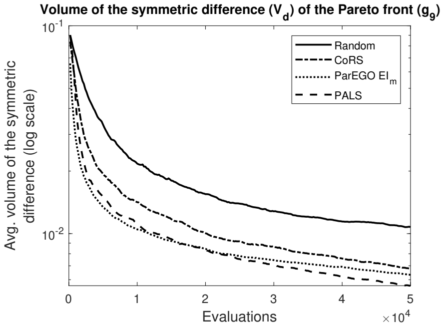


Figure: Average metrics volume of symmetric difference (left) and classification error (right), for a Random approach, an alternative random approach based on probability of non-domination, a scalarization ParEGO adapted with EI_m and PALS, for problem g_9.

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.

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