Bayesian multi-objective optimization for stochastic simulators

Subhasish Basak[1,2], Julien Bect[2] & Emmanuel Vazquez[2]

MASCOT-NUM 2023

April 4, Le Croisic, France

1. Agence Nationale de Sécurité Sanitaire (ANSES), Maison-Alfort, France

2. Université Paris-Saclay, CNRS, CentraleSupeléc, L2S, Gif-sur-yvette, France

This work is part of the ArtiSaneFood project (ANR-18-PRIM-0015) which is part of the PRIMA program supported by EU

Thesis supervisors:

Julien Bect (L2S)

Fanny Tenenhaus-Aziza (CNIEL)

Laurent Guillier (ANSES)

Emmanuel Vazquez (L2S)

Motivation & application

- Project ArtiSaneFood
  - Control risk of illness from food-borne pathogens in cheese
  - Make methodological recommendations to cheese producers
- Quantitative Risk Assessment (QRA) model:
  - Finite input space $\mathbb{X}$:
    - Intervention parameters : Testing frequency, sampling units, etc.
  - Estimates the objectives:
    - Risk of illness $\rightarrow R^{\mathrm{HUS}}$
      conflicting $\uparrow \downarrow$ trade-off
    - Average cost of intervention $\rightarrow C$

Batch simulator : stochastic and expensive

- Based on Perrin et al. (2014) and Basak et al. (in prep.)

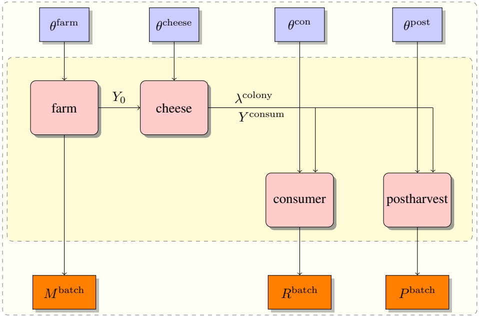


Figure 1: Several batches (milk → cheese) are simulated to estimate R^{HUS} & C

Multiobjective simulation optimization

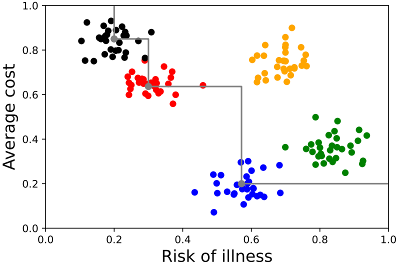


Figure 2: Estimated outputs corresponding to 5 inputs $\mathbb{X} = \{x_1, x_2, x_3, x_4, x_5\}$

- Goal: Identify best trade-offs among conflicting objectives

  → Find optimal inputs that minimize the objectives

Contents

1 Problem formulation

- Consider a biobjective minimization problem of $f = (R^{\mathrm{HUS}}, C)$

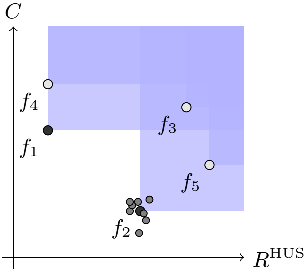
$$\min_{x \in \mathbb{X}} f(x) \tag{1}$$

- Additive noise: $Z(x) = f(x) + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \Sigma)$
- Conflicting objectives: There are no unique optimal solution
- The solution set consists of Pareto optimal points

$$\mathcal{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, f(x') \prec f(x)\} \tag{2}$$

- $f(x') \prec f(x) \implies f_i(x') \leq f_i(x), \forall i$, with at least one strict inequality

Pareto optimal solutions



f_3, f_4 & f_5 dominated by f_1 and f_2

- Goal: Estimate the Pareto set $\mathcal{P}$ and Pareto front $\mathcal{F}$ (image of $\mathcal{P}$)
- Using a limited simulation budget of noisy observations

How to assess performance of an estimate ?

- Volume of symmetric difference on Pareto front $\mathcal{F}$ (image of $\mathcal{P}$)

  $\rightarrow$ Uses dominated region between $\mathcal{F}$ and a reference point R

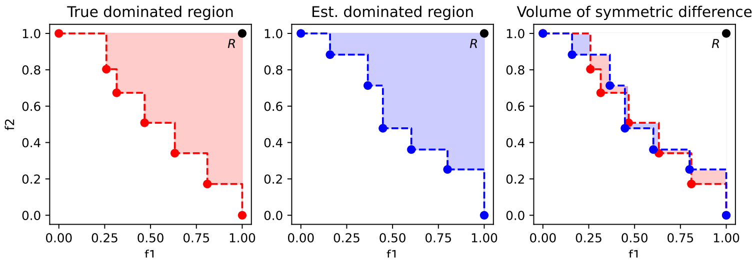


Figure 3: True Pareto front, Estimated Pareto front and reference R

Other performance metrics

- Misclassification rate on Pareto set $\mathcal{P}$

  $\rightarrow$ Uses estimated $\widehat{\mathcal{P}}$ and true $\mathcal{P}^*$ Pareto set

$$M(\mathcal{P}^*, \widehat{\mathcal{P}}) = \frac{1}{|\mathbb{X}|} \sum_{x \in \mathbb{X}} |\mathbb{1}_{x \in \mathcal{P}^*} - \mathbb{1}_{x \in \widehat{\mathcal{P}}}|$$

- Many more:
  - Indicator of correct selection
  - Hausdorff distance
  - ...

Bayesian Optimization (BO) framework

Algorithm 1 BO with a Gaussian process (GP) prior ξ on f

Sample* f at n_0 points
while budget > 0 **do**
  Update : GP posterior ξ_n
  Compute : acquisition function $J_n(x)$
  Optimize : $x_{n+1} = \arg\max_{x \in \mathbb{X}} J_n(x)$
  Sample* : f at x_{n+1}
end while
Estimate** $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{F}}$

$*f$ is sampled with a batch size k
$**$ $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{F}}$ are estimated with GP posterior mean $\mu_n(x)$

Bayesian sequential design of computer experiments

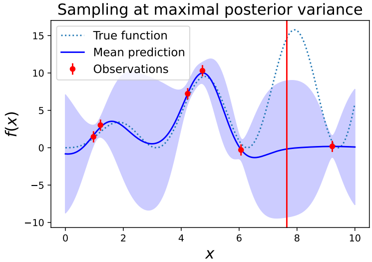
How to construct an acquisition function $J(x)$?

- Two standard branches in the literature:
  - Maximal Uncertainty Sampling (MUS)
  - Stepwise Uncertainty Reduction (SUR)
- Heuristically and experimentally, global uncertainty strategies based on SUR are often found to be more efficient than MUS methods
- Framework: Multiobjective stochastic optimization

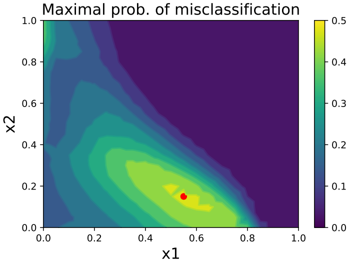
  $\rightarrow$ Study performance of both MUS and SUR methods

2 Maximal uncertainty sampling

- Idea : Sample at $x \in \mathbb{X}$ where uncertainty is maximum
- Application in different design of experiment frameworks:



(a) Function approximation
Sampling at maximal posterior variance



(b) Estimation of probability of failure
Sampling at maximal misclassification prob.

A first idea

- Sampling at maximal probability of misclassification (Bryan et al. (2005))
- $J_n(x)$: Bernoulli variance of the indicator function $\mathbb{1}_{\{x \in \mathcal{P}\}}$

$$\mathrm{Var}_n\left(\mathbb{1}_{\{x \in \mathcal{P}\}}\right) = p_n(x) \cdot (1 - p_n(x))$$

$\rightarrow$ Estimated with conditional simulations of ξ_n (Binois et al. (2015))

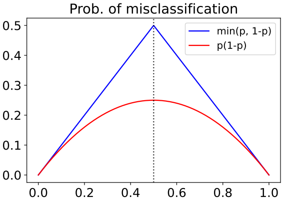
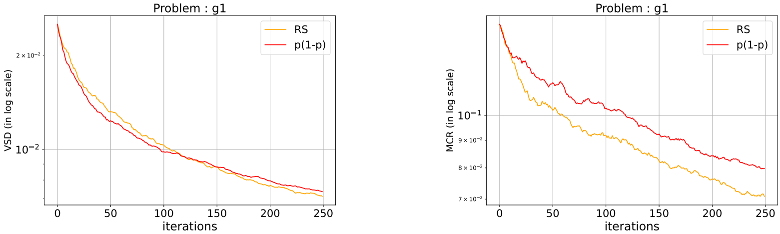


Figure 5: Equivalent measures in literature (Bect et al. (2011))

Performance on test problem g1 from Barracosa et al. (2021)



(a) Vol. of symm. difference

(b) Misclassification rate

Figure 6: Average performance of $p_n \cdot (1 - p_n)$ based MUS method compared to a naive Random Search (RS) baseline method that samples X_{n+1} randomly from $\mathbb{X}$

This method does not work!!

Why it fails ?

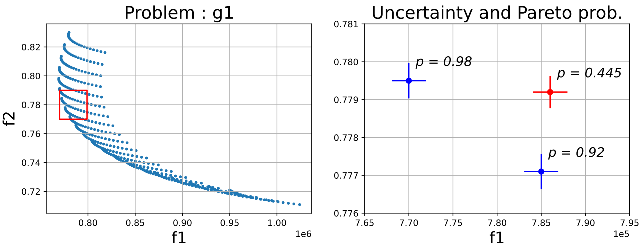


Figure 7: Zoomed area from left figure on right, with $p_n(x)$ and uncertainty

- A point with small uncertainty can have $p_n(x) \approx 0.5$ due to its neighbours
- The algorithm gets stuck at such points and samples repetitively

Weighted Mean Squared Error (w-MSE)

- Not all uncertainty measures can be reduced by MUS
- We try to reduce the weighted mean squared error
- Weights target the "potential Pareto optimal" region

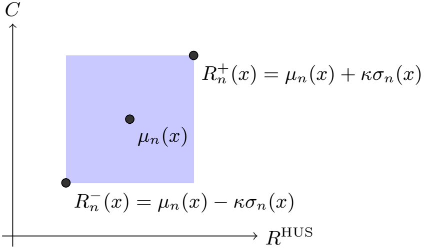
$$X_{n+1} = \arg\max_{x \in \mathbb{X}} \left(w_n(x) \cdot \sum_{j=1}^{q} \frac{\sigma_{j,n}^2(x)}{R_{j,n}^2} \right)$$

$R_{j,n}$ is a normalizing constant for $j = 1, 2, \ldots, q$ -th objective

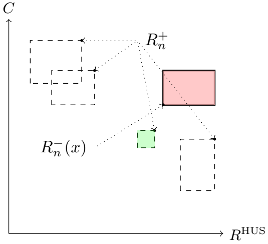
- Choice of weights ?
  - 0-1 weights (corresponds to the PAL algorithm, Zuluaga et al., 2013)
  - Other choices of $w_n(x)$ are tested (see Appendix A)

PALS (PAL + stochastic setting): Barracosa et al. (2021)

- PALS is a w-MSE algorithm with $w_n(x) = \mathbb{1}_{\{x \in \mathbb{X} \setminus N_n\}}$



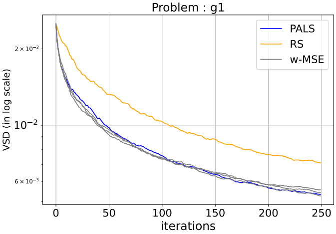
(a) Confidence rectangle



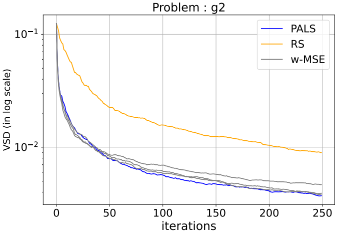
(b) $R_n^+(x') \prec R_n^-(x)$

- $x \in N_n$ if optimistic $R_n^-(x)$ is dominated by some pessimistic $R_n^+(x')$

Performance on test problems from Barracosa et al. (2021)



(a) VSD on g1



(b) VSD on g2

Figure 9: Average performance of PALS with w-MSE methods and Random Search

No consistent improvement over PALS!!

Conclusions & perspectives

- Tested on problems g1-g9 (Barracosa et al. (2021))
- No w-MSE method is consistently better than PALS

Can we do better with SUR?

3 Stepwise uncertainty reduction?

- Origins & applications : reliability theory, optimization, . . .
  - Vazquez and Piera Martinez (2006), Villemonteix et al. (2007), Vazquez and Bect (2009), . . .
- Quantification of uncertainty at step n: H_n
  - Sample the next observation that minimizes the acquisition function

$$J_n(x) = \mathbb{E}_n(H_{n+1}|X_{n+1} = x)$$

  - $\mathbb{E}_n$: conditional expectation w.r.t $\{X_1, Z_1, \ldots, X_n, Z_n\}$
  - $J_n(x)$ does not always have a closed analytical form

Weighted Integrated Mean Squared Error

- w-IMSE : Special case* of SUR method, for some particular choice of $w_n(x)$

$$X_{n+1} = \arg\min_{x \in \mathbb{X}} \sum_{i=1}^{|\mathbb{X}|} w_n(x_i) \sum_{j=1}^{q} \frac{\sigma^2_{j,n+1}(x_i|x)}{R^2_{j,n}}$$

How to choose a good weight $w_n(x)$ function?

- A first idea* : $w_n(x_i) = p_n(x_i)$
- PALS based : $w_n(x_i) = \mathbb{1}_{\{x_i \in \mathbb{X} \setminus N_n\}}$
- Based on uncertainty measures on $\mathcal{F}$ (see, e.g., Binois, 2015)
- Our proposed method : combines PALS classification and measures on $\mathcal{F}$

Proposed method for computing weights

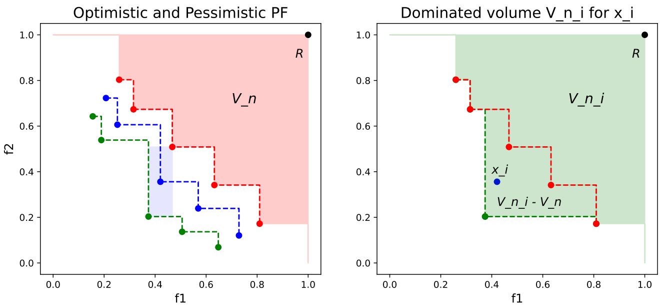


Figure 10: Pareto front $\widehat{\mathcal{F}}$, $\widehat{\mathcal{F}_p}$ and $\widehat{\mathcal{F}_o}$ with dominated volumes V_n and $V_{n,i}$

$$w_n(x_i) \leftarrow \mathbb{1}_{\{x_i \in \mathbb{X} \setminus N_n\}} \cdot (V_{n,i} - V_n)$$

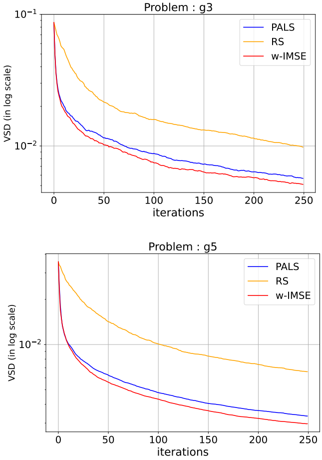


Figure 11: Average VSD metric on problems g-3,4,5,6 (Barracosa et al. (2021))

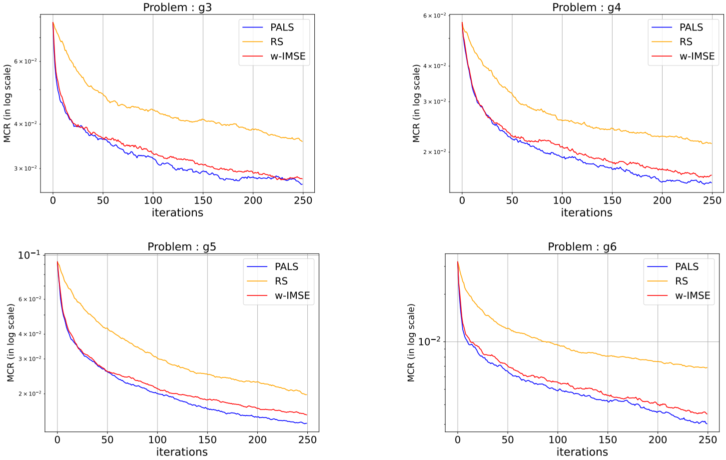


Figure 12: Average MCR metric on problems g-3,4,5,6 (Barracosa et al. (2021))

4 Conclusions & perspectives

- Our proposed w-IMSE method improves consistently over PALS on $\mathcal{F}$
  - The method was designed to improve the VSD metric on $\mathcal{F}$
  - Significant improvement on simulation budget for a given performance
- Trade-off : No other tested methods beat PALS w.r.t both MCR and VSD
- Remarks on PALS
  - A simple, inexpensive, easy-to-implement w-MSE algorithm
  - Not easy to beat even with complicated SUR methods

Open Questions?

- How to improve over PALS on the MCR metric?
  - Tractable SUR strategies over w-IMSE?
  - w-IMSE with weights based on uncertainty on $\mathcal{P}$?
- How to improve over PALS on both MCR and VSD metric?
- ...

Thank you for your attention!

COPIL ArtiSaneFood 2022, Caen, Normandie

References

F. Perrin, F. Tenenhaus-Aziza, V. Michel, S. Miszczycha, N. Bel, and M. Sanaa. Quantitative risk assessment of haemolytic and uremic syndrome linked to O157:H7 and non-O157:H7 shiga-toxin producing escherichia coli strains in raw milk soft cheeses. Risk Analysis, 2014.

S. Basak, J. Christy, L. Guillier, F. Audiat-Perrin, M. Sanaa, F. Tenenhaus-Aziza, J. Bect, and E. Vazquez. Quantitative risk assessment of haemolytic and uremic syndrome (hus) from consumption of raw milk soft cheese. in prep.

B. Bryan, R. C. Nichol, C. R. Genovese, J. Schneider, C. J. Miller, and L. Wasserman. Active learning for identifying function threshold boundaries. In Advances in Neural Information Processing Systems, 2005.

M. Binois, D. Ginsbourger, and O. Roustant. Quantifying uncertainty on pareto fronts with gaussian process conditional simulations. European Journal of Operational Research, 2015.

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing, 2011.

B. Barracosa, J. Bect, H. Dutrieux Baraffe, J. Morin, J. Fournel, and E. Vazquez. Extension of the pareto active learning method to multi-objective optimization for stochastic simulators. Virtual Conference originally scheduled in Fort Worth, Texas, United States, 2021.

M. Zuluaga, G. Sergent, A. Krause, and M. Püschel. Active learning for multi-objective optimization. Proceedings of Machine Learning Research, 2013.

E. Vazquez and M. Piera Martinez. Estimation of the volume of an excursion set of a gaussian process using intrinsic kriging, 2006.

J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions, 2007.

E. Vazquez and J. Bect. A sequential bayesian algorithm to estimate a probability of failure. IFAC Proceedings Volumes, 2009.

M. Binois. Uncertainty quantification on pareto fronts and high-dimensional strategies in bayesian optimization, with applications in multi-objective automotive design. Theses, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2015.

A Appendix

Weights based on dominated area

Algorithm 2 Construction of $m(x)$ at step n

Estimate Pareto set $\widehat{\mathcal{P}_n}$ base on GP posterior mean
$\mathcal{F}_n^+ \leftarrow \{R_n^+(x) | x \in \widehat{\mathcal{P}_n}\}$ ▷ (Pessimistic Pareto Front)
$V_{\text{ref}} \leftarrow D(\mathcal{F}_n^+)$ ▷ (Reference dominated volume)
for $x \in \mathcal{C}$ **do**
    $\mathcal{F}_n \leftarrow$ Pareto front of $\{\mathcal{F}_n^+ \cup R_n^-(x)\}$
    $V_n \leftarrow D(\mathcal{F}_n^+)$ ▷ (Dominated volume)
    $m(x) \leftarrow (V_n - V_{\text{ref}})/V_{\text{ref}}$ ▷ (normalized in $[0, 1]$)
end for
Return $m(x)$

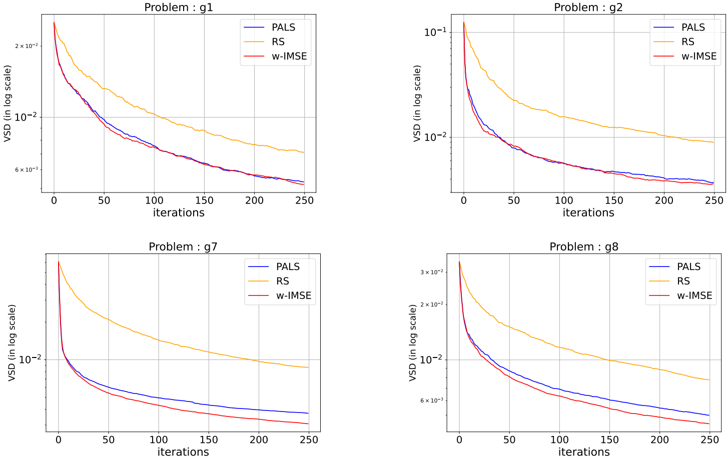


Figure 13: Average VSD metric on g-1,2,7,8 (Barracosa et al. (2021))

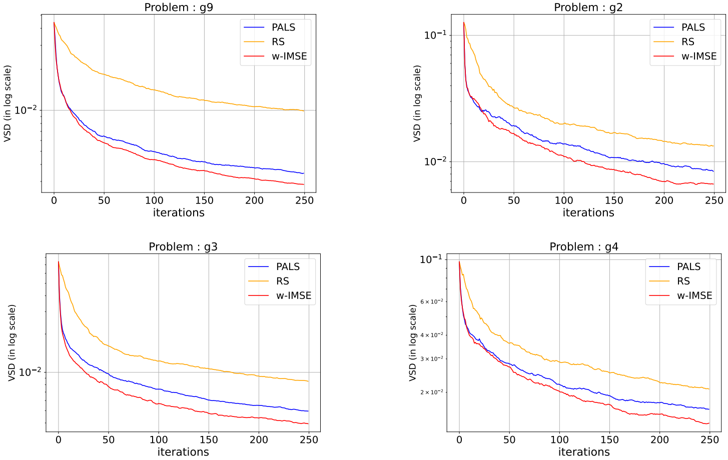


Figure 14: Average VSD metric on g-9 (Barracosa et al. (2021)), g-2,3,4-bis

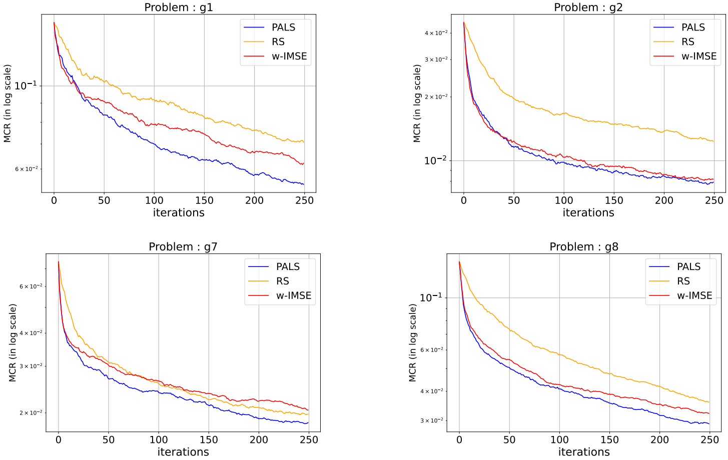


Figure 15: Average MCR metric on g-1,2,7,8 (Barracosa et al. (2021))

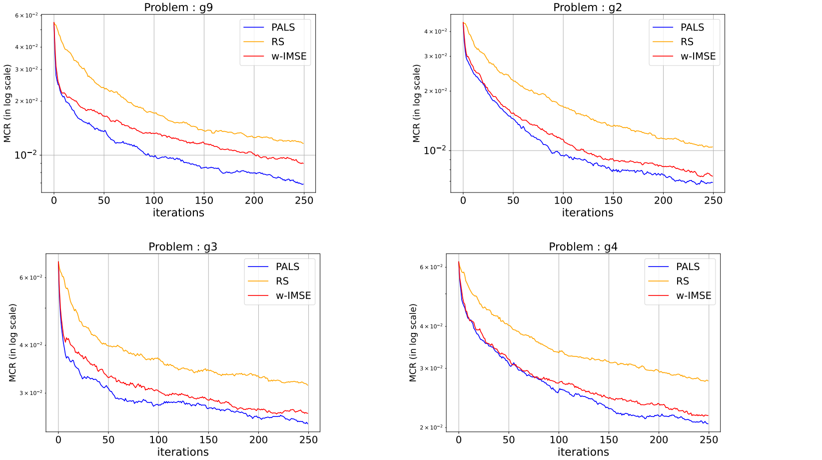


Figure 16: Average MCR metric on g-9 (Barracosa et al. (2021)), g-2,3,4-bis

Weighted-MSE methods

- w-MSE-α
  - $w_n(x) = \mathbb{1}_{\{x \in \Gamma_n^\alpha\}}$

$$\Gamma_n^\alpha = \{x | p_n(x) > \alpha \cdot \min_{x \in \widehat{\mathcal{P}_n}} p_n(x)\}$$

  - Tested on $\alpha = 0.1, 0.5$
- w-MSE-λ
  - $w_n(x) = \lambda \cdot p_n(x) + (1 - \lambda) \cdot p_n(x) \cdot (1 - p_n(x))$
  - Tested on $\lambda = 0, 0.33, 1$

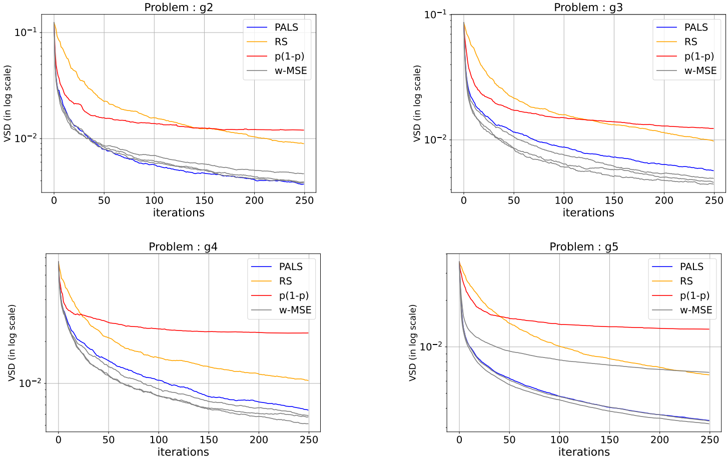


Figure 17: Average VSD metric on g-2,3,4,5 (Barracosa et al. (2021))

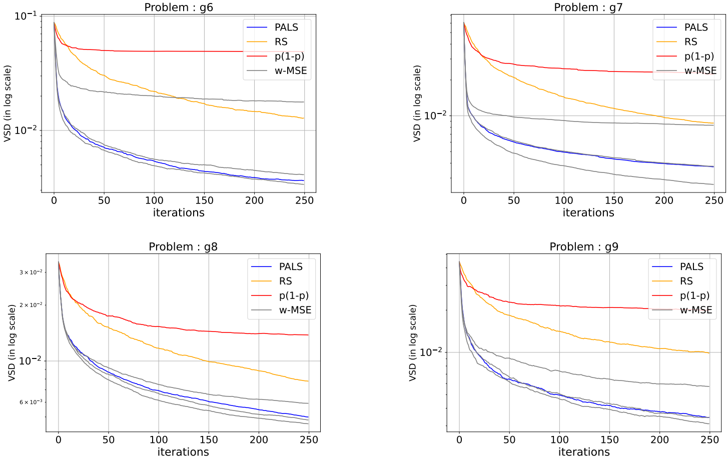


Figure 18: Average VSD metric on g-6,7,8,9 (Barracosa et al. (2021))

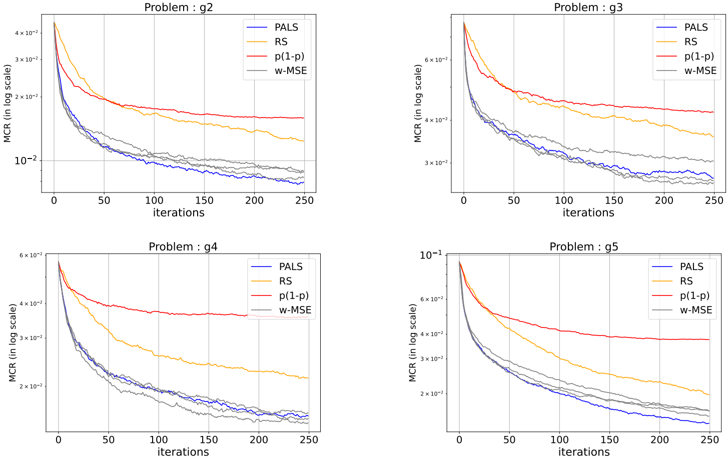


Figure 19: Average MCR metric on g-2,3,4,5 (Barracosa et al. (2021))

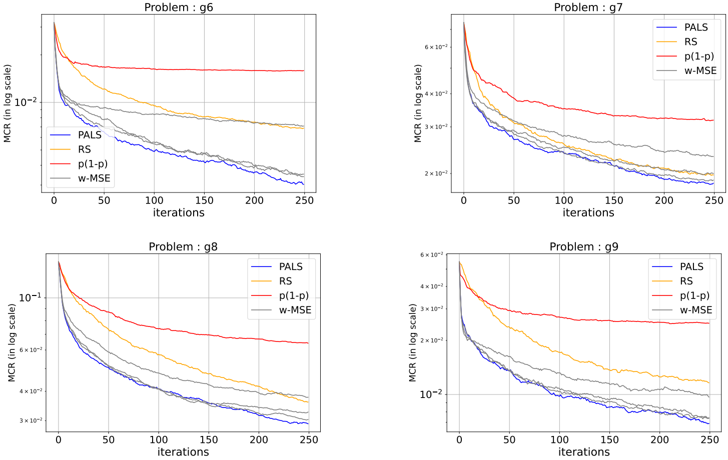


Figure 20: Average MCR metric on g-6,7,8,9 (Barracosa et al. (2021))