Response to Reviewer 1: We thank the reviewer for recognizing the quality, clarity, and thorough experiments.

1. Yes, Algorithm 1 should return both Pareto set (PS) and Pareto front (PF). PS is the set of non-dominated points in $\mathcal{D}$ (data of function evaluations) and can be computed with one pass over $\mathcal{D}$. PF is the function evaluations of points in PS.

2. We compute ideal Pareto front via exhaustive search on open-source data sets. This is possible for all our experiments

Response to Reviewer 2: We thank the reviewer for recognizing the quality, clarity, and thorough experiments.

1. We will add a paragraph in the introduction to clarify the challenges in extending MES to multi-objective (MO) setting. However, please note that we clearly mentioned that this work is an extension of MES to MO setting. The title of the paper/algorithm clearly conveys this information. We repeatedly cited MES (**9 times**) at each step where there is a similarity or a change (once in introduction, once in background/setup, twice in technical section, five times in appendix). We emphasize the challenges in extending MES to MO setting and our proposed solutions below.

2. Although notationally and thematically some steps in MESMO might be similar to MES, there are multiple steps that are not straightforward. We employed mathematical identities and proofs that were neither used in MES nor in PESMO. Equation 4.9 along with its proof by contradiction in the following paragraph is a key step essential for the final expression. This is followed by the use of entropy identity to decompose the expression in equation 4.10. These choices allowed the later use of truncated Gaussian entropy expression, which would have been impossible otherwise. On the other hand, PESMO used expectation propagation to approximate the entropy computation over which MESMO has multiple advantages as described in the related work section.

3. Equations 4.4-4.6 are different from the corresponding **notationally** equivalent equations (1-3) in MES. In MESMO, the information gain (IG) is defined w.r.t $\mathcal{Y}^*$, the true Pareto front (non-dominated multi-dimensional vectors with varying trade-off between objectives). IG in MES is computed with respect to y^*, the single-dimensional true maximum.

4. Theoretical results of MESMO are inspired from MES, but extending them to MO setting required multiple new changes (details are provides in appendix). MO analysis accounted for the trade-off between multiple functions (distribution over a random vector); and also targeted MO-related metric (R2), which is quite different from SO.

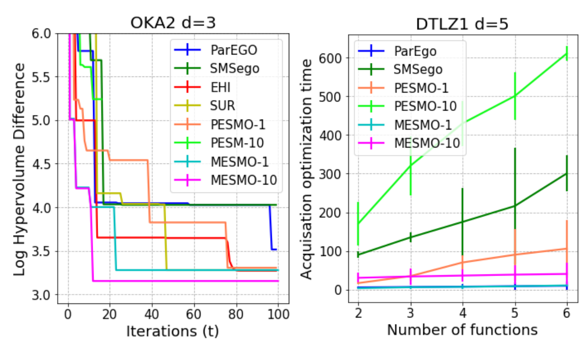
Response to Reviewer 3: We thank the reviewer for recognizing the novelty and significance of the proposed work.

1. We will modify the introduction to mention all related algorithms. We will also clarify the relation to MES and challenges in extending MES to MO setting. We emphasized PESMO in the introduction because it is an input space entropy based method. ParEGO, SMSego, and EIHV were cited in the related work.

2. Related work will be improved by including more details as follows. EHI and SUR becomes very slow as no. of functions increases. The reason SMSego is relatively faster is that the gain in hypervolume is computed over a limited set of points: SMSego finds those set of points by optimizing the posterior means of the GPs.

3. Algorithms that reduce the MO problem to a single-objective using scalarization (similar to Parego), typically choose the scalars randomly. The randomness might lead to sub-optimal results. By sub-optimality, we mean a solution that does not compare well with the true Pareto front (see references [6,14,16,18] in paper).

4. The left figure shows results of experiment run on a benchmark from the general MO literature (OKA2). We have results ready for multiple benchmarks and will include them in the final paper. We opted for building our synthetic benchmarks by combining conflicting functions from the single-objective BO literature as ML community is more familiar with them. We also included results on **four** real-world benchmarks from prior BO papers for MO (PESMO [6]; PAL [26]). The right figure shows the optimization time with a fixed dimension = 5 with increasing number of objective functions.



5. Entropy based methods are a class of non-myopic acquisition functions that consider how evaluating a new point impacts the global understanding of the black box function. They are known to have superior performance in single-objective BO setting. MESMO is based on the same principle and appropriately captures the trade-off between multiple functions leading to superior performance. It should be noted that PESMO is based on the same principle for MO setting, but MESMO has several advantages over PESMO as described in the related work.

6. One potential limitation of MESMO (shared by PESMO) is that computation time increases with the no. of Pareto front samples in each iteration. However, our experiments show that MESMO performs very well even with one sample.

7. We provided details for the open-source Spearmint repository used for SMSego and EHI. The repository calls the *pygmo* library to compute the HV. According to their documentation, the algorithm used is from *Nowak K, Märtens M, Izzo D., Empirical performance of the approximation of the least hypervolume contributor, PPSN 2014, pp 662-671.*