# LIKELIHOOD-FREE MULTI-OBJECTIVE BAYESIAN OPTIMIZATION UNDER HETEROSCEDASTIC NOISE

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Multi-objective Bayesian optimization (MOBO) fits a probabilistic surrogate, typically one Gaussian process (GP) per objective, so every acquisition function inherits the noise model of that surrogate. When the noise is heteroscedastic, that model cannot be identified from a single evaluation per design, and both the search and its evaluation degrade. We propose likelihood-free acquisition functions that do not require a noise model. We define a soft non-dominance indicator by nonparametric subsampling of the noisy observed objective vectors, derive it in closed form, and use it to weight a classification objective, which yields the likelihood-free counterparts of two state-of-the-art noisy acquisitions. These acquisition functions depend on the data only through the dominance relations between observations, and we prove that they preserve dominance and that their maximizers eventually lie in the true Pareto set. We further show that the standard evaluation protocols do not penalize a method for the set it would recommend, and we replace it by an exact decomposition of the decision maker's loss into sampling, set-selection and mis-selection terms, of which we observe mis-selection accounts for more than $40\%$ at high noise. Across a comprehensive experimental campaign, GP-based baselines remain competitive when the noise is low and well specified, but as the noise level and structural complexity grow our acquisitions outperform GP baselines by margins that widen with the input dimension.

## 1 INTRODUCTION

Many design problems in science and engineering trade off several expensive black-box objectives whose evaluations are noisy. For example, an expensive stochastic simulator returns a different output at every run, or a physical experiment is subject to measurement error and to conditions that cannot be controlled. For deterministic objectives the field is mature, with exact methods, scalarization and evolutionary heuristics (Emmerich & Deutz, 2018; Li et al., 2018) assessed by set indicators whose properties are understood (Zitzler et al., 2008; Auger et al., 2012). Under noise little of that carries over, and the gap matters: what a decision maker needs is not only a good search, but a guarantee on the set an algorithm *returns*, namely that its designs are the ones whose expected performance is truly non-dominated. Noisy multi-objective black-box optimization, and the heteroscedastic case in particular, is still in its infancy by comparison.

Multi-objective Bayesian optimization (MOBO) addresses such problems by fitting a probabilistic surrogate, usually one Gaussian process (GP) per objective or a multi-output GP, and maximizing an acquisition function that scores a candidate by what its evaluation is worth to the Pareto front; e.g., the expected hypervolume improvement (EHVI) (Emmerich et al., 2006; Daulton et al., 2020), random scalarizations (Knowles, 2006; Paria et al., 2020; De Ath et al., 2022), or information-theoretic criteria (Hernández-Lobato et al., 2016; Belakaria et al., 2019; Tu et al., 2022). Methods built explicitly for noisy observations are far scarcer. qNEHVI (Daulton et al., 2021) and its numerically robust successor qLogNEHVI (Ament et al., 2023) integrate over the uncertainty in the incumbent Pareto front, and SK-MORS (Rojas Gonzalez et al., 2020) searches with stochastic kriging models that separate simulation noise from surrogate uncertainty. In all of these the noise enters through the likelihood of the surrogate, so an acquisition is only as reliable as the noise model it is computed on.

Noise in practice is rarely homoscedastic and Gaussian: its magnitude often scales with the response, it can depend on inputs the optimizer neither observes nor controls, and it can contain outliers.



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Heteroscedastic and robust regression is well developed, from stochastic kriging on replicated designs (Ankenman et al., 2010) to latent log-noise processes (Lázaro-Gredilla & Titsias, 2011; Binois et al., 2018), heteroscedastic neural regressors (Immer et al., 2023) and outlier-aware likelihoods (Jylänki et al., 2011; Ament et al., 2024), and it has been used in single-objective BO (Griffiths et al., 2022; Makarova et al., 2021). In MOBO, such a model must be fit for each of the $m$ objectives. In many cases, replications are not available, and we must learn this model from a *single* evaluation per design, a task that becomes harder as the input dimension grows, where even the mean function is difficult to learn from a small budget (Binois & Wycoff, 2022; Hvarfner et al., 2024). The effect of heteroscedastic noise on MOBO has not been studied systematically.

Performance evaluation is also a problem in its own right: finding promising designs does not guarantee that a method can reliably identify which ones to recommend. For deterministic objective vectors, a *Pareto-compliant* indicator ensures that a set dominating another cannot receive a worse score (Zitzler et al., 2007; 2008). Under noise, however, the objective vectors must be estimated, so these indicators cannot guarantee the correct rank of the sets consistently with their true (unknown) performance (Basseur & Zitzler, 2006; Branke, 2025). Comparisons based on predictive distributions or pairwise dominance tests likewise depend on inference from noisy data and can misclassify dominance relationships. The prevailing evaluation practice sidesteps this uncertainty by scoring the true objective vectors of every evaluated design and using an oracle to select the truly nondominated subset (Daulton et al., 2021; Ament et al., 2023). This measures a method's ability to sample good designs, but not its ability to recommend them from the noisy observations available to it.

We treat the identification of non-dominated designs as classification with noisy labels. Likelihood-free Bayesian optimization (LFBO) (Song et al., 2022), which builds on density-ratio estimation (Tiao et al., 2021; Oliveira et al., 2022), obtains an acquisition function from a classifier trained on utility-weighted labels, without modeling the objectives. In the case of noise-free non-dominance labels, this classifier estimates the probability of hypervolume improvement (Steinberg et al., 2025). Noise corrupts these labels at input-dependent rates, and standard label-noise corrections require exactly those rates (Natarajan et al., 2013), which we cannot identify with one evaluation per design. We therefore soften the labels by subsampling the evaluated set without replacement in the spirit of stability selection (Meinshausen & Bühlmann, 2010), and record how often each design remains non-dominated. The resulting *soft indicator* responds to noisy near-ties and to designs dominated by a single competitor. Because it depends on the observations only through their dominance relations, it needs no noise assumption and is invariant to monotone transformations of each objective. Appendix A relates this view to simulation-optimization and pure-exploration bandits literature, which control the same misclassification risk through repeated measurements or evaluations.

**Contributions.** We derive the soft indicator in closed form (Lemma 1) and build from it LF-NPHVI and LF-NEHVI, likelihood-free counterparts of noisy PHVI and EHVI that need neither a noise model nor per-objective surrogates. LF-NEHVI weights the indicator by the hypervolume improvement over each subsampled front, and so marginalizes over the incumbent front as qNEHVI does (Section 3). We prove that the subsampled utilities decay geometrically in the number of observations dominating a design (Lemma 2), and that the population acquisition separates strictly dominated designs from their dominators, so every population maximizer is Pareto optimal (Proposition 2; both in Appendix C.2). To evaluate the methods, we score each method's recommended Pareto set with two estimators applied identically to every method: the subsampling estimator and a variational heteroscedastic GP (Lázaro-Gredilla & Titsias, 2011). We prove that the oracle, returned-set and decision-maker conventions are ordered and split the decision maker's loss exactly into sampling, set selection and mis-selection (Proposition 3); at the highest noise level, mis-selection accounts for more than $40\%$ of this loss (Section 4). Finally, experiments over 16 benchmarks (two to five objectives, 5 to 20 inputs, four noise structures, three noise levels) show GP-based methods ahead at low noise and likelihood-free methods ahead from moderate noise onward, with an advantage that grows with input dimension (Section 5).

## 2 Background and Problem Setting

**Noisy multi-objective optimization.** We maximize $m \geq 2$ black-box objectives $\boldsymbol{f} = (f_1, \ldots, f_m) \colon \mathcal{X} \to \mathbb{R}^m$ over a box $\mathcal{X} \subset \mathbb{R}^d$. Evaluating a design $\boldsymbol{x}$ returns one noisy observation

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}(\boldsymbol{x}), \qquad \mathbb{E}[\boldsymbol{\varepsilon}(\boldsymbol{x}) \mid \boldsymbol{x}] = \mathbf{0}, \tag{1}$$

where the distribution of $\varepsilon(\boldsymbol{x})$ may depend on $\boldsymbol{x}$ and need not be Gaussian. For two objective vectors $\boldsymbol{z}, \boldsymbol{z}' \in \mathbb{R}^m$, $\boldsymbol{z}$ *dominates* $\boldsymbol{z}'$, written $\boldsymbol{z} \succ \boldsymbol{z}'$, if $z_k \geq z'_k$ for all $k$ and $z_k > z'_k$ for some $k$. The target is the Pareto set of the expected objectives, $\mathcal{P}^\star = \{\boldsymbol{x} \in \mathcal{X} : \nexists\, \boldsymbol{x}' \in \mathcal{X} \text{ with } \boldsymbol{f}(\boldsymbol{x}') \succ \boldsymbol{f}(\boldsymbol{x})\}$, and its image $\boldsymbol{f}(\mathcal{P}^\star)$, the Pareto front. After $n$ sequential evaluations the data are $\mathcal{D}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^n$, and

$$\mathcal{P}_n = \{\, i \leq n : \nexists\, j \leq n \text{ with } \boldsymbol{f}(\boldsymbol{x}_j) \succ \boldsymbol{f}(\boldsymbol{x}_i) \,\} \tag{2}$$

indexes the designs that are truly non-dominated among those evaluated. We assume a design is evaluated once, so neither $\boldsymbol{f}(\boldsymbol{x}_i)$ nor the noise variance at $\boldsymbol{x}_i$ can be estimated from replications, and $\mathcal{P}_n$ is known neither to the optimizer nor to the decision maker. Estimating $\mathcal{P}_n$ from $\{\boldsymbol{y}_i\}_{i \leq n}$ can incur two *misclassification* errors: a design in $\mathcal{P}_n$ can appear dominated (*false negative* or *exclusion*), and a design outside $\mathcal{P}_n$ can appear non-dominated (*false positive* or *inclusion*).

**Hypervolume-based acquisition.** For a finite set $A \subset \mathbb{R}^m$ and a reference point $\boldsymbol{r}$, the hypervolume $\mathrm{HV}(A) = \lambda_m(\bigcup_{\boldsymbol{a} \in A}[\boldsymbol{r}, \boldsymbol{a}])$ is the Lebesgue measure of the region that $A$ dominates above $\boldsymbol{r}$, and the hypervolume improvement of $\boldsymbol{z}$ is $\mathrm{HVI}(\boldsymbol{z}; A) = \mathrm{HV}(A \cup \{\boldsymbol{z}\}) - \mathrm{HV}(A)$. It is positive if and only if $\boldsymbol{z}$ exceeds $\boldsymbol{r}$ in every objective and no element of $A$ is at least as good as $\boldsymbol{z}$ in every objective. GP-based MOBO typically places a GP prior on each $f_k$ and scores a candidate by a posterior functional of HVI against the current front, such as its expectation (EHVI) or the probability that it is positive (PHVI) (Emmerich et al., 2006; Daulton et al., 2020). Under noise the current front is itself uncertain, and qNEHVI (Daulton et al., 2021) averages HVI over joint posterior samples of $\boldsymbol{f}$ at the candidate and at all evaluated designs; qLogNEHVI (Ament et al., 2023) optimizes a smoothed logarithm of the same quantity. In every case the noise is represented by the Gaussian likelihood of the GPs. The original evaluation of qNEHVI supplies the true noise variances to the model (Daulton et al., 2021). We do not give any method this information: each GP baseline infers one noise variance per objective by maximum a posteriori estimation under `BoTorch's` default priors (Balandat et al., 2020), which is what a practitioner without noise measurements would do.

**Likelihood-free Bayesian optimization.** LFBO (Song et al., 2022) replaces the surrogate by a probabilistic classifier $\pi_\theta \colon \mathcal{X} \to (0, 1)$. Given a non-negative utility $u_i$ for each evaluated design, it minimizes the weighted cross-entropy

$$\mathcal{L}(\theta) = -\frac{1}{n} \sum_{i=1}^{n} [u_i \log \pi_\theta(\boldsymbol{x}_i) + \log(1 - \pi_\theta(\boldsymbol{x}_i))], \tag{3}$$

in which every design enters as a negative example of unit weight and as a positive example of weight $u_i$. If $u_i$ is a noisy realization with conditional mean $\bar{u}(\boldsymbol{x}) = \mathbb{E}[u \mid \boldsymbol{x}]$, minimizing the population version of (3) pointwise gives

$$\frac{\pi^\star(\boldsymbol{x})}{1 - \pi^\star(\boldsymbol{x})} = \bar{u}(\boldsymbol{x}), \tag{4}$$

so the odds of the trained classifier rank designs by expected utility and serve as the acquisition function. Rescaling all $u_i$ by a common constant, for instance to balance the two classes, rescales the odds and leaves their maximizer unchanged. With $u = \mathbb{1}[y > \tau]$ or $u = (y - \tau)^+$ (where $(\cdot)^+ = \max\{\cdot, 0\}$) for a threshold $\tau$, (4) recovers the probability and the expectation of improvement without a model of $y$. For multiple objectives, Steinberg et al. (2025) use non-dominance to label each design, and since non-dominance is equivalent to $\mathrm{HVI} > 0$, they show the resulting classifier estimates PHVI. The utility only has to be computable from the data, so the construction makes no assumption about the noise. Under noise, however, the labels are computed from $\boldsymbol{y}_i$ rather than $\boldsymbol{f}(\boldsymbol{x}_i)$ and carry both inclusion and exclusion errors. Section 3 constructs utilities that account for this.

**Noise structures.** Table 1 lists the four noise structures we study. They share $\mathbb{E}[\boldsymbol{y} \mid \boldsymbol{x}] = \boldsymbol{f}(\boldsymbol{x})$ but differ in whether the variance depends on $\boldsymbol{x}$ and in the shape of the marginal. In the hidden-input structure one input $\xi$ is drawn at random and withheld from the optimizer, which then searches over the remaining $d - 1$ inputs, and the target $f_k$ is the average of the original objective $F_k$ over $\xi$. To make noise levels comparable across structures and problems, each is calibrated to a common *noise factor*

$$\eta = \frac{\sqrt{\mathbb{E}_{\boldsymbol{x}}[\mathbb{V}[\varepsilon_k(\boldsymbol{x}) \mid \boldsymbol{x}]]}}{r_k}, \qquad k = 1, \ldots, m, \tag{5}$$

the root mean noise variance relative to the range $r_k$ of $f_k$ over the search space, with $\boldsymbol{x}$ uniform on that space. We use $\eta \in \{0.05, 0.2, 0.4\}$; the calibration of each structure is given in Appendix B.

Table 1: Noise structures per objective $k$ with $z \sim \mathcal{N}(0,1)$. The constants $\sigma_k$ and $a_k$ are set by (5).

| Structure | Observation $y_k$ | Variance | Marginal |
|---|---|---|---|
| Constant | $f_k(\boldsymbol{x}) + \sigma_k z$ | constant | Gaussian |
| Contaminated | $f_k(\boldsymbol{x}) + s\,\sigma_k z$, $s = 5$ w.p. 0.1, else 1 | constant | heavy-tailed |
| Multiplicative | $f_k(\boldsymbol{x}) \exp(\sigma_k z - \sigma_k^2/2)$ | $\propto f_k(\boldsymbol{x})^2$ | log-normal, skewed |
| Hidden input | $f_k(\boldsymbol{x}) + a_k[F_k(\boldsymbol{x}, \xi) - f_k(\boldsymbol{x})]$, $\xi$ unobserved | depends on $\boldsymbol{x}$ | induced by $F_k$ |

## 3 Proposed acquisition functions under heteroscedastic noise

The classifier of Section 2 needs one non-negative number per evaluated design. Hard non-dominance labels are a poor choice under noise: a single lucky observation is enough to flip the label. We therefore replace the labels by quantities that measure how *stable* a design's non-dominance is, and we obtain them by perturbing the set of competitors rather than by modeling the noise.

### 3.1 A soft non-dominance indicator

Write $Y_n = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$ for the observed objective vectors and $\mathrm{ND}(\cdot)$ for the non-dominated subset of a finite set. Fix a subsample fraction $\gamma \in (0,1)$ and let $s = \lfloor \gamma n \rfloor$. Draw $B$ index sets $I^{(1)}, \dots, I^{(B)}$ of size $s$ uniformly without replacement, and let $P^{(b)} = \mathrm{ND}(\{\boldsymbol{y}_j\}_{j \in I^{(b)}})$ be the Pareto front of subsample $b$. The *soft non-dominance indicator* of design $i$ is the fraction of subsamples in which it survives,

$$\zeta_i = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\left[\nexists\, \boldsymbol{p} \in P^{(b)} \text{ with } \boldsymbol{p} \succ \boldsymbol{y}_i\right] \in [0,1]. \tag{6}$$

Subsampling is done without replacement on purpose. A bootstrap would duplicate observations, and since identical vectors never dominate each other, duplicates would inflate every front and bias $\zeta$ upward. The half-sample $\gamma = 1/2$ is the default of stability selection (Meinshausen & Bühlmann, 2010), which perturbs the comparison set as strongly as possible while keeping enough competitors for the dominance structure to mean something.

Although (6) is defined by sampling, its $B \to \infty$ limit is available in closed form and depends on the data only through the number of observations that dominate $\boldsymbol{y}_i$.

**Lemma 1** (Exact soft non-dominance). *Let $k_i = |\{\, j \le n : \boldsymbol{y}_j \succ \boldsymbol{y}_i \,\}|$. As $B \to \infty$,*

$$\zeta_i \;\longrightarrow\; \frac{\binom{n-k_i}{s}}{\binom{n}{s}} \;=\; \prod_{t=0}^{k_i-1} \frac{n-s-t}{n-t}, \tag{7}$$

*with the convention that the product equals 1 for $k_i = 0$ and $\zeta_i = 0$ when $n - k_i < s$.*

The proof is in Appendix C.1: a front point of the subsample dominates $\boldsymbol{y}_i$ if and only if the subsample contains at least one of the $k_i$ observations that dominate $\boldsymbol{y}_i$, so (6) is a hypergeometric probability. Three properties make $\zeta$ suitable under unknown noise. (1) *Graded evidence:* $\zeta_i$ decreases with the number of observations dominating $\boldsymbol{y}_i$: at $\gamma = 1/2$, a design dominated once keeps $\zeta_i = \lceil n/2 \rceil / n \approx 1/2$, while one dominated twice falls below $1/4$. Exclusion errors, typically caused by a single lucky competitor, are thus softened, whereas a design dominated by many observations is labeled negative regardless of the noise. (2) *No noise model:* $\zeta_i$ depends on $Y_n$ only through dominance relations, so it is invariant to strictly increasing transformations of each objective; heavy tails, skew and heteroscedasticity matter only through the comparisons they flip. And (3) *finite-budget uncertainty:* Removing competitors also probes how much of a design's apparent non-dominance stems from competitors not yet evaluated.

The fronts $\{P^{(b)}\}_{b \le B}$ form an ensemble of plausible Pareto fronts drawn from the data, and $\zeta$ is the frequency with which a design belongs to them. In random-set terms, $\zeta$ estimates the coverage function of this ensemble (Molchanov, 2005; Binois et al., 2015), a nonparametric counterpart of the posterior over fronts that a GP would provide.

**Algorithm 1** One round of LF-NPHVI / LF-NEHVI

**Require:** data $\mathcal{D}_n = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i \le n}$, subsample fraction $\gamma$, subsamples $B$, reference $\boldsymbol{r}$ (LF-NEHVI)
1: $s \leftarrow \lfloor \gamma n \rfloor$
2: **for** $b = 1, \dots, B$ **do**
3: draw $I^{(b)} \subset \{1, \dots, n\}$ of size $s$ uniformly without replacement; $P^{(b)} \leftarrow \mathrm{ND}(\{\boldsymbol{y}_j\}_{j \in I^{(b)}})$
4: score every design against $P^{(b)}$: $\zeta_i^{(b)}$ in (6) for LF-NPHVI, or $\nu_i^{(b)}$ in (9) for LF-NEHVI
5: **end for**
6: $u_i \leftarrow$ average of the $B$ scores ▷ $\zeta_i$, resp. $\nu_i$
7: fit $\pi_\theta$ by minimizing (3) with weights $u_i$ rescaled to mean one
8: **return** $\boldsymbol{x}_{n+1} \in \arg\max_{\boldsymbol{x} \in \mathcal{X}} \log(\pi_\theta(\boldsymbol{x})/(1 - \pi_\theta(\boldsymbol{x})))$

### 3.2 LF-NPHVI

Taking $u_i = \zeta_i$ in the LFBO loss (3) defines **LF-NPHVI**. By (4) the trained classifier satisfies

$$\frac{\pi^\star(\boldsymbol{x})}{1 - \pi^\star(\boldsymbol{x})} = \mathbb{E}[\zeta \mid \boldsymbol{x}], \tag{8}$$

the probability that a design at $\boldsymbol{x}$ would not be dominated by the front of a random subsample of the evaluated data. Since non-dominance with respect to a front is the same event as positive HVI against it, (8) is a probability of HVI in which the incumbent front is random rather than fixed. Without noise and with $\gamma \to 1$ the indicator collapses to the hard label and (8) recovers the PHVI of Steinberg et al. (2025). Computing $\zeta$ needs no hypervolume, so LF-NPHVI carries no cost that grows with the number of objectives beyond the dominance tests themselves.

### 3.3 LF-NEHVI

A probability says whether a design is likely to reach the front, not by how much it would improve it. To recover the magnitude we weight each subsample by hypervolume. For subsample $b$ with front $P^{(b)}$, define

$$\nu_i^{(b)} = \begin{cases} \mathrm{HVI}\big(\boldsymbol{y}_i;\, P^{(b)}\big) & \text{if } \boldsymbol{y}_i \notin P^{(b)}, \\ \mathrm{HV}\big(P^{(b)}\big) - \mathrm{HV}\big(P^{(b)} \setminus \{\boldsymbol{y}_i\}\big) & \text{if } \boldsymbol{y}_i \in P^{(b)}, \end{cases} \qquad \nu_i = \frac{1}{B} \sum_{b=1}^{B} \nu_i^{(b)}. \tag{9}$$

The first case is the usual hypervolume improvement, and it is zero for every design that the subsampled front dominates. The second case is needed because a design that is already on $P^{(b)}$ improves it by nothing, so the plain definition would give the current best designs no weight at all; its exclusive contribution, the hypervolume that would be lost if it were removed, is the natural leave-one-out replacement and keeps the two cases on the same scale. Taking $u_i = \nu_i$ in (3) defines **LF-NEHVI**, whose classifier odds estimate the EHVI of a design averaged over the ensemble of subsampled fronts. This is the same marginalization over an uncertain incumbent front that qNEHVI performs with posterior samples (Daulton et al., 2021), obtained here from the data alone. The weights enter (3) only up to a common factor, so the scale of the HV does not affect the acquisition.

### 3.4 The algorithm and its cost

Algorithm 1 summarizes a round and Algorithm 2 (Appendix D) the outer BO loop. Both acquisitions replace the $m$ GP fits of a round, at a cost independent of the noise distribution; since only LF-NEHVI computes hypervolumes, we use it for moderate numbers of objectives ($m \le 5$ here). We fix $\gamma = 1/2$ and $B = 128$ as untuned defaults: $\gamma = 1/2$ is the half-sample of Section 3.1, and by Lemma 1, $B$ only controls the Monte Carlo error around a closed-form limit. Appendix F reports the cost of a round and ablates both parameters, where terminal quality is flat in $B$ and changes little with $\gamma$.

The classifier is an ensemble of five anchored multilayer perceptrons with three residual blocks each, each regularized toward its own random initialization so that the members disagree away from the data (Pearce et al., 2020; Oliveira, 2026). The acquisition is the mean logit, maximized by multi-start gradient ascent (details in Appendix D). Any classifier trained with (3) can replace the ensemble in Algorithm 1.

**Pareto alignment.** For a theoretical analysis of the acquisitions independently of classifier approximation, Appendix C.2 considers their population counterparts $a_n(\boldsymbol{x}) := \mathbb{E}[u(\boldsymbol{y}; P \mid \boldsymbol{x}, \mathcal{D}_{n-1}]$, where $\boldsymbol{y}$ is a fresh observation at $\boldsymbol{x}$ and $P$ is a random subsampled Pareto front as in Section 3.1. Proposition 2 shows that, under the stated conditions, at every iteration,

$$\boldsymbol{f}(\boldsymbol{x}') \preceq \boldsymbol{f}(\boldsymbol{x}), \quad \boldsymbol{f}(\boldsymbol{x}') \neq \boldsymbol{f}(\boldsymbol{x}) \quad \Longrightarrow \quad a_n(\boldsymbol{x}') < a_n(\boldsymbol{x}). \tag{10}$$

Hence every maximizer of the population acquisition lies in the true Pareto set. The result applies to continuous design spaces; controlling the gap between the population acquisition and its fitted classifier approximation remains a direction for future work.

## 4 Evaluating noisy MOBO

Noisy MOBO is conventionally scored on the true objectives (i.e., the "oracle") of every evaluated design. Writing $\mathcal{A}$ for a fixed reference Pareto front of the problem and $d^+$ for the modified distance of Ishibuchi et al. (2015), with all objectives divided by the extent of $\mathcal{A}$, the reported quantities are

$$\mathrm{HV}_n^{\mathrm{orc}} = \mathrm{HV}(\{\boldsymbol{f}(\boldsymbol{x}_i)\}_{i \le n}), \qquad \mathrm{IGD}^{+,\mathrm{orc}}_n = \frac{1}{|\mathcal{A}|} \sum_{\boldsymbol{a} \in \mathcal{A}} \min_{i \le n} d^+(\boldsymbol{a}, \boldsymbol{f}(\boldsymbol{x}_i)). \tag{11}$$

Both evaluate every design in the set at its noise-free objective values, so the non-dominated designs are selected using information the optimizer never has. Without noise this is the appropriate measure. Under noise, however, a method can sample the front well and still return the wrong designs, because the decision maker must choose among designs based on noisy information. We therefore score the set each method recommends, and keep (11) as an oracle reference.

### 4.1 Pareto-set estimators.

An estimator $E$ maps the data $\mathcal{D}_n$ to two quantities per evaluated design: a membership score $\hat{\pi}_i^E \in [0, 1]$, the estimated degree to which design $i$ belongs to $\mathcal{P}_n$, and an objective estimate $\hat{\boldsymbol{y}}_i^E \in \mathbb{R}^m$, the objective vector the decision maker believes design $i$ has. The *returned set* is

$$\widehat{\mathcal{S}}_n^E = \{\, i \le n : \hat{\pi}_i^E \ge \tfrac{1}{2} \,\}, \tag{12}$$

falling back to the single highest-scoring design if this set is empty. The threshold is parameter-free and, for a calibrated membership probability, minimizes the expected number of misclassified designs. Both estimators below are applied identically to the data of every method: they use no noise variances, no acquisition and no information of the method that produced the data, so differences between methods reflect the designs they evaluated rather than the quality of their own surrogates. This also makes methods without a surrogate comparable to the GP baselines.

**Likelihood-free estimator.** Scores membership with the exact soft indicator of Lemma 1 at $\gamma = 1/2$, the same notion the proposed acquisitions are built on, and takes the observation itself as the objective estimate: $\hat{\pi}_i^{\mathrm{LF}} = \zeta_i$ and $\hat{\boldsymbol{y}}_i^{\mathrm{LF}} = \boldsymbol{y}_i$. By (7), $\zeta_i \ge 1/2$ at $\gamma = 1/2$ exactly when $k_i \le 1$, so

$$\widehat{\mathcal{S}}_n^{\mathrm{LF}} = \{\, i \le n : \text{at most one observation dominates } \boldsymbol{y}_i \,\} \tag{13}$$

always contains the observed non-dominated designs. It is agnostic about the noise and fits nothing, but it cannot pool information across designs, so the decision maker selects among raw observations.

**Variational heteroscedastic GP estimator.** This is the GP-based counterpart of (13). Each objective $k$ is modeled independently as

$$y_k(\boldsymbol{x}) = g_k(\boldsymbol{x}) + \varepsilon_k(\boldsymbol{x}), \qquad \varepsilon_k(\boldsymbol{x}) \sim \mathcal{N}\big(0,\, e^{h_k(\boldsymbol{x})}\big), \qquad g_k \sim \mathcal{GP}(0, \kappa_g), \quad h_k \sim \mathcal{GP}(\mu_0, \kappa_h), \tag{14}$$

with $g_k$ and $h_k$ independent a priori. The latent log-noise process $h_k$ models input-dependent variance. Both processes are fitted by maximizing the variational bound of Lázaro-Gredilla & Titsias (2011); a flat $h_k$ recovers a homoscedastic GP, so that case lies inside the model family (App. D). Writing $\boldsymbol{g} = (g_1, \ldots, g_m)$ and taking $S = 1024$ joint posterior draws $\boldsymbol{g}^{(s)}$ at the evaluated designs,

$$\hat{\boldsymbol{y}}_i^{\mathrm{VH}} = \mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_i) \mid \mathcal{D}_n], \qquad \hat{\pi}_i^{\mathrm{VH}} = \frac{1}{S} \sum_{s=1}^{S} \mathbb{1}\big[\, \nexists\, j \le n \text{ with } \boldsymbol{g}^{(s)}(\boldsymbol{x}_j) \succ \boldsymbol{g}^{(s)}(\boldsymbol{x}_i) \,\big] \tag{15}$$

estimates the posterior probability that design $i$ is non-dominated among those evaluated. This estimator pools across designs, represents input-dependent variance, and shares no component with any likelihood-free method. A homoscedastic GP would instead impose the GP baselines' own noise assumption on the evaluation, but three of the four noise structures violate it by construction.

**Metrics on the returned set.** We score $\widehat{\mathcal{S}} = \widehat{\mathcal{S}}_n^E$ on the true objectives with hypervolume and with IGD$^+$, that is, (11) restricted to $i \in \widehat{\mathcal{S}}$, and with the noisy IGD$^+$ of Branke (2025): for every target $\boldsymbol{a} \in \mathcal{A}$ the decision maker picks the design in $\widehat{\mathcal{S}}$ whose *estimated* objectives are closest to the target, and is scored on its *true* objectives,

$$\mathrm{nIGD}^{+,E}_n = \frac{1}{|\mathcal{A}|} \sum_{\boldsymbol{a} \in \mathcal{A}} d^+\big(\boldsymbol{a}, \boldsymbol{f}(\boldsymbol{x}_{j^E(\boldsymbol{a})})\big), \qquad j^E(\boldsymbol{a}) = \arg\min_{i \in \widehat{\mathcal{S}}} \left\| \widetilde{\boldsymbol{a}} - \widetilde{\hat{\boldsymbol{y}}}_i^E \right\|_2, \tag{16}$$

where a tilde denotes normalization by the extent of $\mathcal{A}$. The three indicators answer different questions: what the returned designs are worth as a set, and what the decision maker gets when it also has to pick one of them.

**The loss decomposition.** The three quantities are ordered, for every run, checkpoint and estimator: $\mathrm{IGD}^{+,\mathrm{orc}}_n \le \mathrm{IGD}^{+,E}_n \le \mathrm{nIGD}^{+,E}_n$ and $\mathrm{HV}^E_n \le \mathrm{HV}^{\mathrm{orc}}_n$, because the returned set indexes a subset of the evaluated designs and the decision maker's pick is one of its members. The chain looks into an exact accounting of what that decision maker loses,

$$\mathrm{nIGD}^{+,E}_n = \underbrace{\mathrm{IGD}^{+,\mathrm{orc}}_n}_{\text{sampling}} + \underbrace{\big(\mathrm{IGD}^{+,E}_n - \mathrm{IGD}^{+,\mathrm{orc}}_n\big)}_{\text{set selection}} + \underbrace{\big(\mathrm{nIGD}^{+,E}_n - \mathrm{IGD}^{+,E}_n\big)}_{\text{mis-selection}}, \tag{17}$$

with all three terms non-negative. The sampling term is estimator-free: it is the best the evaluated designs could support. The set-selection term is what the decision maker loses because the returned set leaves out designs that matter; it vanishes exactly when that set contains every truly non-dominated evaluated design, in which case the hypervolume is unchanged as well. The mis-selection term is the loss from picking within the set by estimated rather than true objectives, which dominated designs in the set inflate. The oracle convention reports the first term and charges nothing for the other two. Proposition 3 in Appendix C.3 states the ordering and the equality condition formally and proves them, and Section 5 quantifies how much the oracle convention leaves out.

## 5 Experiments

We use a combination of well-known test functions (Huband et al., 2006) and real-life examples in two input-dimension groups: low-dimensional real-life examples include VehicleSafety, Penicillin and CarSideImpact ($m = 3, 4$ and $d = 5$ to 7). Higher-dimensional test-functions include DTLZ1 and DTLZ2 at $d = 15$ and $d = 20$ with $m = 2, 3$, DTLZ2 and DTLZ7 with $m = 5$ ($d = 14$), WFG6 with $m = 4$ ($d = 16$), and ZDT1 and ZDT2 ($d = 20$, $m = 2$). WFG6 is taken from `pymoo` (Blank & Deb, 2020) and all other problems from `BoTorch` (Balandat et al., 2020), the framework we use to set up our experimental campaign. For performance evaluation, we compute hypervolumes exactly with `moocore`[1], which we measured to be two to four orders of magnitude faster than the implementation currently used in `BoTorch`, from about $10^2$ times at $m = 2$ to more than $4 \times 10^4$ times at $m = 5$ (see Appendix E for more details). A run evaluates $2(d + 1)$ Sobol designs and then 100 designs one at a time, with a single noisy evaluation per design. Every GP-based baseline fits one GP per objective and infers its own noise. Appendix D gives the full settings. We report terminal HV gap, IGD$^+$ and nIGD$^+$ under the three estimators of Section 4; since nIGD$^+$ requires an estimate of the objectives, it is undefined for the oracle.

**Baselines** Sobol is the standard non-adaptive reference. qLogEHVI (Ament et al., 2023) maximizes EHVI against the observed front, treating it as if it were noise-free, and measures what ignoring the noise costs. qLogNParEGO draws a random scalarization each round (Knowles, 2006; Daulton et al., 2021), which avoids hypervolume computations. qLogNEHVI (Daulton et al., 2021; Ament et al., 2023) is the noise-aware member of the family, integrating over the posterior of the incumbent front, which is precisely what the proposed acquisitions replace, so Table 6 in Appendix G compares specifically against it. Finally, qLB-JES (Tu et al., 2022) represents the state of the art in information-theoretic acquisitions; we use its `BoTorch` implementation without tuning (Appendix D).

[1] https://github.com/multi-objective/moocore

Table 2: Mean rank of the seven methods at the final iteration (lower is better; best per row and block in bold), taken per condition over the seed mean, on all problem instances considered. The three blocks score the same runs under the likelihood-free estimator, the VHGP estimator and the oracle values (11); nIGD$^+$ needs an estimate of the objectives so is not defined for the oracle.

| | | likelihood-free estimator | | | | | | | VHGP estimator | | | | | | | oracle values | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| metric | $\eta$ | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
| $\log_{10}$ HV gap | 0.05 | 6.73 | 2.38 | **1.98** | 4.28 | 5.64 | 2.97 | 4.01 | 6.61 | 2.67 | **2.13** | 4.40 | 5.48 | 2.78 | 3.92 | 6.73 | 2.50 | **1.87** | 4.30 | 5.71 | 2.97 | 3.92 |
| | 0.2 | 6.64 | 4.17 | 3.80 | 4.39 | 4.00 | **2.25** | 2.75 | 6.49 | 4.27 | 3.83 | 4.33 | 3.91 | **2.32** | 2.87 | 6.76 | 4.28 | 3.59 | 4.39 | 4.17 | **2.05** | 2.77 |
| | 0.4 | 6.40 | 4.12 | 4.02 | 4.29 | 4.47 | 2.40 | **2.30** | 6.03 | 4.08 | 4.14 | 4.45 | 4.45 | **2.37** | 2.49 | 6.69 | 4.17 | 4.03 | 4.22 | 4.52 | **2.15** | 2.23 |
| IGD$^+$ | 0.05 | 6.39 | 2.66 | **2.31** | 2.73 | 5.41 | 3.52 | 4.98 | 6.23 | 2.66 | **2.31** | 3.41 | 5.48 | 3.30 | 4.61 | 6.41 | 2.70 | **2.19** | 2.77 | 5.48 | 3.48 | 4.97 |
| | 0.2 | 6.36 | 4.38 | 3.75 | 3.59 | 3.50 | **2.83** | 3.59 | 6.06 | 4.41 | 3.92 | 3.94 | 3.44 | **2.75** | 3.48 | 6.67 | 4.30 | 3.61 | 3.66 | 3.52 | **2.61** | 3.64 |
| | 0.4 | 6.00 | 3.88 | 3.77 | 4.08 | 3.97 | 3.17 | **3.14** | 5.31 | 3.92 | 4.23 | 4.33 | 3.81 | **3.17** | 3.22 | 6.52 | 4.09 | 3.80 | 4.05 | 3.81 | **2.81** | 2.92 |
| nIGD$^+$ | 0.05 | 6.36 | 2.83 | 2.53 | **2.42** | 5.30 | 3.69 | 4.88 | 6.00 | 2.73 | **2.36** | 3.47 | 5.53 | 3.44 | 4.47 | not defined | | | | | | |
| | 0.2 | 5.97 | 4.28 | 3.98 | 3.66 | 4.00 | **2.95** | 3.16 | 5.45 | 4.28 | 3.81 | 3.91 | 3.66 | **3.36** | 3.53 | not defined | | | | | | |
| | 0.4 | 5.59 | 3.61 | 3.64 | 3.86 | 4.78 | 3.38 | **3.14** | 5.23 | 3.56 | 3.88 | 3.91 | 3.89 | 4.09 | **3.44** | not defined | | | | | | |

### 5.1 Results and discussion

Table 2 summarizes the campaign as the mean rank of each method at the final iteration, per metric, noise level and estimator. Ranks are computed per condition (problem, noise structure and noise level) from the mean over ten seeds. Table 6 in Appendix G.1 counts wins, ties and losses specifically against qLogNEHVI by means of two-sided Mann–Whitney $U$ tests over seeds at the $5\%$ level: a method wins or loses a condition only when the test rejects, and ties otherwise. The per-problem terminal values in Appendix G.6 mark every method not significantly different from the best. Appendix G.2 adds convergence curves on twenty-four conditions under each estimator.

**The noise level and structure are decisive.** Table 2 shows a clean crossover. At $\eta = 0.05$ qLogNEHVI ranks first on HV gap and IGD$^+$ and the likelihood-free methods are mid-field, as expected of a well-specified GP under small noise. From $\eta = 0.2$ onward the ordering reverses: LF-NPHVI and LF-NEHVI outperform on all metrics, while qLogNEHVI falls to mid-field, close to its noise-unaware counterpart qLogEHVI. The reversal appears under all three estimators; it is therefore a property of where each method samples, not of how its returned set is scored. The estimators differ only in the margins: under VHGP, which smooths the objectives before selecting, the lead on nIGD$^+$ narrows, and at $\eta = 0.4$ only LF-NEHVI stays ahead of the GP baselines. Split by noise structure (Appendix G.3), the crossover on HV gap appears under all four, homoscedastic Gaussian included, so it follows the amount of noise rather than misspecification alone.

**The advantage grows with input dimension.** Against qLogNEHVI (Table 6 and Appendix G.2) on high-dimensional problems the proposed methods converge orders of magnitude faster. This result survives a change of estimator, with wins/ties/losses of 24/20/4 under VHGP, which shares nothing with the proposed methods, and 34/10/4 under the oracle values, which measure sampling alone. On nIGD$^+$ under VHGP the comparisons are mostly ties: given a smoothing model, both families return sets of similar decision quality and differ mainly in where they sampled. At lowest $\eta$ values and low-dimensionality, qLogNEHVI is the strongest on average, but as $\eta$ increases, the low-dimensional group turns as well: the proposed methods perform at least as good as qLogNEHVI.

**Convergence.** Figure 1 shows one example condition per metric at $\eta = 0.4$. Where the likelihood-free methods lead, they separate within the first ten to twenty iterations, before a GP has enough data to fit a noise level, and the lead then holds or widens. We also note that the two proposed acquisitions do not always move together, even though they differ only in how the subsampled front is weighted.

**Half of the loss is identification, not sampling.** Under the likelihood-free estimator, the mis-selection share of the decision maker's loss (17) grows from $0.26$ at $\eta = 0.05$ to $0.42$ at $\eta = 0.4$, and the set-selection share from $0.02$ to $0.12$, so at the highest level barely half of the loss is the sampling term that the oracle convention reports. The VHGP estimator reaches the same verdict differently: selecting on smoothed estimates keeps its mis-selection share near $0.30$, but its returned set shrinks from $1.10$ to $0.50$ times the size of $\mathcal{P}_n$ as the posterior widens, and its set-selection share grows to

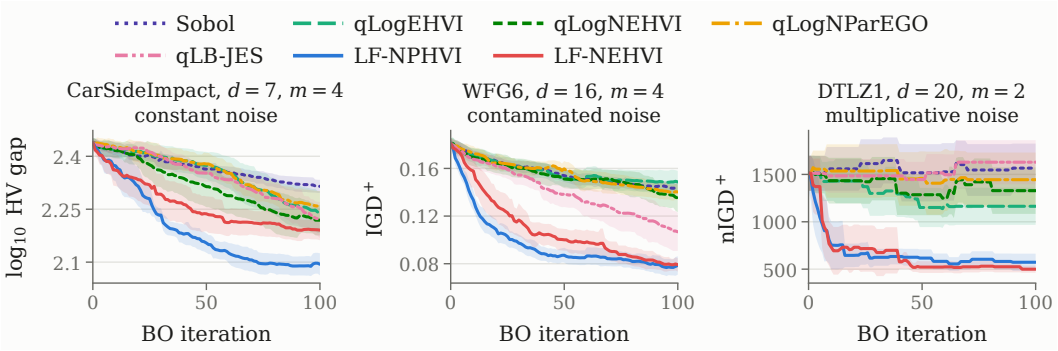


Figure 1: Three example conditions at the highest noise level ($\eta = 0.4$), one per metric, as mean over ten seeds with a $\pm 2$ standard error band on the returned set of the likelihood-free estimator. Appendix G.2 shows twenty-four further conditions under each of the three estimators, with the terminal values of every condition in Appendix G.6.

0.22 instead (Table 10, Appendix G.4). Identification is the larger share on the low-dimensional problems, where the evaluated designs already sit close to the front (Table 11, Appendix G.5). The ordering of Proposition 3 holds numerically in all 13,432 runs.

**Cost.** The median cost per iteration separates the two families as $m$ grows (Table 4, Appendix F): qLogNEHVI rises from 4.5 seconds at $m = 2$ to 192.9 at $m = 5$, while LF-NPHVI, which never computes a hypervolume, stays between 3.3 and 7.8 seconds. LF-NEHVI, which does, takes 46.0 seconds at $m = 4$, and qLB-JES grows from 8.1 to 28.3 seconds.

**Why the distance-based indicators struggle under noise.** $\mathrm{IGD}^+(S)$ (11) is weakly Pareto-compliant (Ishibuchi et al., 2015), but the noisy protocol scores $\widehat{\mathcal{S}}_n$, and $\widehat{\mathcal{S}}_n \not\subseteq \widehat{\mathcal{S}}_{n+1}$ because a later observation can dominate a design the estimator had kept. The scored sets are thus ordered neither by inclusion nor by dominance, and the indicator inherits no monotonicity in the budget. $\mathrm{nIGD}^+$ further replaces the minimum by an index $j(\boldsymbol{a})$ chosen from the estimates $\hat{\boldsymbol{y}}$, which removes the envelope altogether: a truly dominated design can win the nearest-neighbor assignment of many targets and be charged for all of them, leaving only the bound $\mathrm{nIGD}^+ \geq \mathrm{IGD}^+$ of Proposition 3.

**Limitations.** Benchmarking under uncertainty still lacks agreed indicators, noise structures and test problems, and building that standard would help this area more than another acquisition function. Our results reinforce it: once the scored set is itself estimated, Pareto compliance is lost and a better set need not score better, so our metrics are decision-quality functionals rather than compliant indicators. This shows in the results: 72% of the $\mathrm{nIGD}^+$ curves end above their best value, compared with 36% for HV gap, so in some panels every method appears to get worse as data accumulates (Figure 3, Appendix G.2); and at $\eta = 0.05$ the two estimators disagree on the top-ranked method.

## 6 Conclusion

We take on the challenge to make multi-objective Bayesian optimization work when the noise is neither small, white, Gaussian nor independent from the inputs, and the approach we propose avoids the noise model rather than improving it. A soft non-dominance indicator, obtained by subsampling the evaluated designs and recording how often each survives, turns the acquisition into a classification problem that depends on the data only through the dominance relations between observations. The two acquisitions proposed this way are competitive when the noise is small and clearly outperforming from moderate noise onward, with a margin that grows with the input dimension. This holds for all four noise structures, for three ways of scoring the result, and against the strongest available baselines.

The second half of the argument is about performance evaluation. Scoring every evaluated design through an oracle reports only where a method sampled; scoring the set a method recommends splits the decision maker's loss into sampling, set selection and mis-selection, and at the highest noise level more than 40% of it is the mis-selection that the oracle convention never charges. That accounting is what noisy multi-objective benchmarks have been missing; giving it a guarantee, in the way Pareto compliance underwrites the deterministic case, is the natural next step.

### AI USE STATEMENT

In this work, we used generative AI tools to review the mathematical formulation of the likelihood-free methods, to help formulate and check the proofs of the theoretical results, to assist with the software implementation of the proposed methods and the experimental pipeline, and to help find related work during background research. We also used generative AI tools to review and refine the research ideas, although the original concepts are the authors' own. We have not used generative AI tools to produce or alter experimental data or results; all data and results were generated by the code included in the supplementary material. Additionally, we used generative AI tools to help structure the manuscript, for proofreading, and for critical review of the text. We have reviewed all AI-assisted work. All mathematical statements and proofs were checked line by line by the authors; AI-assisted code was reviewed and tested by the authors; and every reference suggested by an AI tool was read and verified by the authors before being cited. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

### REPRODUCIBILITY STATEMENT

The proposed acquisition functions are defined in Section 3, with the per-round procedure in Algorithm 1 and the full optimization loop in Algorithm 2 (Appendix D). For the theoretical results, the standing assumptions and complete proofs are given in Appendix C: Lemma 1 is proved in Appendix C.1, the dominance properties of the subsampled acquisitions in Appendix C.2, and Proposition 3 in Appendix C.3. The benchmark problems, all available in BoTorch or pymoo, and the evaluation protocol are described in Sections 4 and 5; the noise structures and their calibration are specified in Appendix B; and the exact hypervolume computation is discussed in Appendix E. Appendix D lists the initial designs, evaluation budget, seeding, reference points and fronts, and all hyperparameters of the likelihood-free methods, the GP baselines and the VHGP estimator, none of which are tuned per problem; the ablation settings are given in Appendix F. *To support a thorough assessment, the appendices report the complete set of results rather than only aggregates*: per-problem terminal values for every metric, noise level and Pareto-set estimator (Appendix G.6), together with the rank, win/tie/loss, loss-decomposition and cost tables behind the summaries in the main text, so that every claim can be traced to the underlying numbers. *Anonymized source code implementing the methods, baselines and evaluation, including the scripts that produce every figure and table, is provided in the supplementary material*. Upon acceptance, we will release the code as open source, together with the raw data of all optimization runs and the complete set of computed metrics.

## REFERENCES

Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for bayesian optimization. In *Advances in Neural Information Processing Systems 36*, 2023.

Sebastian Ament, Elizabeth Santorella, David Eriksson, Ben Letham, Maximilian Balandat, and Eytan Bakshy. Robust Gaussian processes via relevance pursuit. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Bruce Ankenman, Barry L Nelson, and Jeremy Staum. Stochastic kriging for simulation metamodeling. *Operations Research*, 58(2):371–382, 2010.

Eric A. Applegate, Guy Feldman, Susan R. Hunter, and Raghu Pasupathy. Multi-objective ranking and selection: Optimal sampling laws and tractable approximations via score. *Journal of Simulation*, 14(1):21–40, 2020.

Çağın Ararat and Cem Tekin. Vector optimization with stochastic bandit feedback. In *International Conference on Artificial Intelligence and Statistics*, pp. 2165–2190. PMLR, 2023.

Peter Auer, Chao-Kai Chiang, Ronald Ortner, and Madalina Drugan. Pareto front identification from stochastic bandit feedback. In *Artificial Intelligence and Statistics*, pp. 939–947. PMLR, 2016.

Anne Auger, Johannes Bader, Dimo Brockhoff, and Eckart Zitzler. Hypervolume-based multiobjective optimization: Theoretical foundations and practical implications. *Theoretical Computer Science*, 425:75–103, 2012.

Maximilian Balandat, Brian Karrer, Daniel Jiang, Samuel Daulton, Ben Letham, Andrew G Wilson, and Eytan Bakshy. Botorch: A framework for efficient monte-carlo bayesian optimization. *Advances in neural information processing systems*, 33:21524–21538, 2020.

Bruno Barracosa, Julien Bect, Héloïse Dutrieux Baraffe, Juliette Morin, Josselin Fournel, and Emmanuel Vazquez. Bayesian multi-objective optimization for stochastic simulators: an extension of the pareto active learning method. *arXiv preprint arXiv:2207.03842*, 2022.

Matthieu Basseur and Eckart Zitzler. A preliminary study on handling uncertainty in indicator-based multiobjective optimization. In *Applications of Evolutionary Computing*, pp. 727–739. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-33238-1.

Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Max-value entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems 32*, 2019.

Mickael Binois and Nathan Wycoff. A survey on high-dimensional gaussian process modeling with application to bayesian optimization. *ACM Transactions on Evolutionary Learning and Optimization*, 2(2):1–26, 2022.

Mickaël Binois, David Ginsbourger, and Olivier Roustant. Quantifying uncertainty on pareto fronts with gaussian process conditional simulations. *European Journal of Operational Research*, 243(2): 386–394, 2015.

Mickael Binois, Robert B Gramacy, and Mike Ludkovski. Practical heteroscedastic gaussian process modeling for large simulation experiments. *Journal of Computational and Graphical Statistics*, 27 (4):808–821, 2018.

Julian Blank and Kalyanmoy Deb. pymoo: Multi-objective optimization in python. *IEEE Access*, 8: 89497–89509, 2020.

Juergen Branke. Performance metrics for multi-objective optimisation under noise. *IEEE Transactions on Evolutionary Computation*, 29(4):1449–1455, 2025. doi: 10.1109/TEVC.2024.3438115.

Juergen Branke and Wen Zhang. Identifying efficient solutions via simulation: myopic multi-objective budget allocation for the bi-objective case. *OR Spectrum*, 41(3):831–865, 2019.

Chun-Hung Chen and Loo Hay Lee. *Stochastic Simulation Optimization*, volume 1. World Scientific, Singapore, 2010. doi: 10.1142/7437.

Rajan Filomeno Coelho. Probabilistic dominance in multiobjective reliability-based optimization: theory and implementation. *IEEE Transactions on evolutionary computation*, 19(2):214–224, 2014.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. *Advances in neural information processing systems*, 33:9851–9864, 2020.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Parallel bayesian optimization of multiple noisy objectives with expected hypervolume improvement. *Advances in neural information processing systems*, 34:2187–2200, 2021.

George De Ath, Tinkle Chugh, and Alma AM Rahat. Mbore: multi-objective bayesian optimisation by density-ratio estimation. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 776–785, 2022.

Michael TM Emmerich and André H Deutz. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing*, 17(3):585–609, 2018.

Michael TM Emmerich, Kyriakos C Giannakoglou, and Boris Naujoks. Single-and multiobjective evolutionary optimization assisted by gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439, 2006.

J. E. Fieldsend and R. M. Everson. Multi-objective optimisation in the presence of uncertainty. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pp. 243–250, 2005.

Carlos M. Fonseca, Luís Paquete, and Manuel López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. In *Proceedings of the 2006 Congress on Evolutionary Computation (CEC 2006)*, 2006.

Ryan-Rhys Griffiths, Alexander A. Aldrick, Miguel Garcia-Ortegon, Vidhi Lalchand, and Alpha A. Lee. Achieving robustness to aleatoric uncertainty with heteroscedastic Bayesian optimisation. *Machine Learning: Science and Technology*, 3(1):015004, 2022.

Andreia P. Guerreiro and Carlos M. Fonseca. Computing and updating hypervolume contributions in up to four dimensions. *IEEE Transactions on Evolutionary Computation*, 22(3):449–463, 2018.

Joshua Q Hale, Helin Zhu, and Enlu Zhou. Domination measure: a new metric for solving multiobjective optimization. *INFORMS Journal on Computing*, 32(3):565–581, 2020.

Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah, and Ryan Adams. Predictive entropy search for multi-objective bayesian optimization. In *International conference on machine learning*, pp. 1492–1501. PMLR, 2016.

Simon Huband, Philip Hingston, Luigi Barone, and Lyndon While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10 (5):477–506, 2006.

Susan R. Hunter, Eric A. Applegate, Viplove Arora, Bryan Chong, Kyle Cooper, Oscar Rincón-Guevara, and Carolina Vivas-Valencia. An introduction to multiobjective simulation optimization. *ACM Trans. Model. Comput. Simul.*, 29(1):7:1–7:36, January 2019. ISSN 1049-3301. doi: 10.1145/3299872.

Carl Hvarfner, Erik Orm Hellsten, and Luigi Nardi. Vanilla bayesian optimization performs great in high dimensions. In *Proceedings of the 41st International Conference on Machine Learning*, 2024.

Alexander Immer, Emanuele Palumbo, Alexander Marx, and Julia Vogt. Effective bayesian heteroscedastic regression with deep neural networks. *Advances in Neural Information Processing Systems*, 36:53996–54019, 2023.

Hisao Ishibuchi, Hiroyuki Masuda, Yuki Tanigaki, and Yusuke Nojima. Modified distance calculation in generational distance and inverted generational distance. In *Evolutionary Multi-Criterion Optimization (EMO)*, volume 9019 of *Lecture Notes in Computer Science*, pp. 110–125. Springer, 2015.

Pasi Jylänki, Jarno Vanhatalo, and Aki Vehtari. Robust Gaussian process regression with a Student-t likelihood. *Journal of Machine Learning Research*, 12:3227–3257, 2011.

Efe Mert Karagözlü, Yaşar Cahit Yıldırım, Çağın Ararat, and Cem Tekin. Learning the pareto set under incomplete preferences: pure exploration in vector bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 3070–3078. PMLR, 2024.

Joshua Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*, 10(1): 50–66, 2006.

Cyrille Kone, Marc Jourdan, and Emilie Kaufmann. Pareto set identification with posterior sampling. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1000–1008. PMLR, 2025a.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Constrained pareto set identification with bandit feedback. In *Forty-second International Conference on Machine Learning*, 2025b.

Miguel Lázaro-Gredilla and Michalis K Titsias. Variational heteroscedastic gaussian process regression. In *ICML*, pp. 841–848, 2011.

Loo Hay Lee, Ek Peng Chew, Suyan Teng, and David Goldsman. Finding the non-dominated pareto set for multi-objective simulation models. *IIE Transactions*, 42(9):656–674, 2010.

Kaiwen Li, Rui Wang, Tao Zhang, and Hisao Ishibuchi. Evolutionary many-objective optimization: A comparative study of the state-of-the-art. *IEEE Access*, 6:26194–26214, 2018.

Anastasia Makarova, Ilnura Usmanova, Ilija Bogunovic, and Andreas Krause. Risk-averse heteroscedastic Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

Ilya Molchanov. *Theory of random sets*. Springer, 2005.

Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Rafael Oliveira. Kernel-based guarantees for nonlinear parametric models in bayesian optimization. *arXiv preprint arXiv:2605.13160*, 2026.

Rafael Oliveira, Louis Tiao, and Fabio T Ramos. Batch Bayesian optimisation via density-ratio estimation with guarantees. *Advances in Neural Information Processing Systems (NeurIPS)*, 35: 29816–29829, 2022.

Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. A flexible framework for multi-objective bayesian optimization using random scalarizations. In *Uncertainty in Artificial Intelligence*, pp. 766–776. PMLR, 2020.

Tim Pearce, Felix Leibfried, and Alexandra Brintrup. Uncertainty in neural networks: Approximately bayesian ensembling. In *International conference on Artificial Intelligence and Statistics*, pp. 234–244. PMLR, 2020.

Sebastian Rojas Gonzalez and Inneke Van Nieuwenhuyse. A survey on kriging-based infill algorithms for multiobjective simulation optimization. *Computers & Operations Research*, 116:104869, 2020.

Sebastian Rojas Gonzalez, Hamed Jalali, and Inneke Van Nieuwenhuyse. A multiobjective stochastic simulation optimization algorithm. *European Journal of Operational Research*, 284(1):212 – 226, 2020.

Sebastian Rojas Gonzalez, Ivo Couckuyt, and Joshua Knowles. Bi-objective stochastic simulation optimization on integer lattices via scalarization. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 511–519, 2026.

Jiaming Song, Lantao Yu, Willie Neiswanger, and Stefano Ermon. A general recipe for likelihood-free Bayesian optimization. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 20384–20404. PMLR, 17–23 Jul 2022.

Daniel M. Steinberg, Asiri Wijesinghe, Rafael Oliveira, Piotr Koniusz, Cheng Soon Ong, and Edwin V. Bonilla. Amortized Active Generation of Pareto Sets. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Louis C Tiao, Aaron Klein, Matthias W Seeger, Edwin V Bonilla, Cedric Archambeau, and Fabio Ramos. Bore: Bayesian optimization by density-ratio estimation. In *International Conference on Machine Learning*, pp. 10289–10300. PMLR, 2021.

Ben Tu, Axel Gandy, Nikolas Kantas, and Behrang Shafei. Joint entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Rick van der Laag, Hao Wang, Thomas Bäck, and Yali Fan. Center-outward $q$-dominance: a sample-computable proxy for strong stochastic dominance in stochastic multi-objective optimisation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26037–26045, 2026.

Eckart Zitzler, Dimo Brockhoff, and Lothar Thiele. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In *Evolutionary multi-criterion optimization*, pp. 862–876. Springer, 2007.

Eckart Zitzler, Joshua Knowles, and Lothar Thiele. Quality assessment of pareto set approximations. *Multiobjective Optimization*, pp. 373–404, 2008.

Marcela Zuluaga, Andreas Krause, and Markus Püschel. e-pal: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research*, 17(104):1–32, 2016.

## A Related work

**Dominance under noise, and misclassification.** The problem of stochastic multi-objective optimization has also been studied by the simulation-optimization community, which offers a complementary view (Chen & Lee, 2010; Hunter et al., 2019). Under noise, dominance itself becomes a modeling choice: one may compare expected values (Rojas Gonzalez & Van Nieuwenhuyse, 2020), the probability that one random objective vector dominates another (Fieldsend & Everson, 2005; Coelho, 2014), or the measure of the set of designs that dominate a given one (Hale et al., 2020), and the Pareto front becomes a random set whose uncertainty can be quantified (Binois et al., 2015). Hale et al. (2020) turn the last of these into an objective in its own right: the domination measure of a design is the probability mass of the designs that dominate it under a sampling distribution, which is a scalar, estimable by sampling, and minimized exactly on the Pareto set. van der Laag et al. (2026) instead strengthen the relation itself, defining a center-outward $q$-dominance from optimal-transport quantile contours that implies strong first-order stochastic dominance between the objective distributions. Both keep the comparison distributional, but need many samples per design. Our indicator gives up the distributional statement in exchange for a single evaluation per design, since it reads only the dominance relations among the observations.

The central risk in this view is misclassification: a Pareto-optimal design that appears dominated (*misclassification by exclusion*, a false negative), or a dominated design that appears Pareto optimal (*misclassification by inclusion*, a false positive). The literature divides according to how these errors are controlled. Search procedures control them implicitly, for instance by classifying Pareto membership from the prediction intervals of a GP or another parametric surrogate whose noise is estimated along the way (Zuluaga et al., 2016; Barracosa et al., 2022). *Multi-objective ranking and selection* procedures control them explicitly, allocating replications over a finite set of candidates, such as the set an algorithm returns (Lee et al., 2010; Applegate et al., 2020; Branke & Zhang, 2019; Rojas Gonzalez et al., 2026); they identify rather than search. Both presume an evaluation budget that MOBO typically does not have: each design is evaluated once and, as argued in Section 1, a reliable noise model for every objective is out of reach.

**Pareto set identification with bandit feedback.** The same misclassification risk is studied in the pure-exploration bandit literature, where the design space is a finite set of arms and each pull returns a noisy vector reward. The setting is the one faced by multi-objective bandits and reinforcement learning, in which every pull or episode returns a noisy vector reward and the learner must decide which arms or policies are non-dominated before spending its budget on them. Auer et al. (2016) give an algorithm that identifies the Pareto front of $K$ such arms with a fixed confidence, together with a sample-complexity bound that scales with problem-dependent dominance gaps. Later work sharpens the sampling rule and the stopping rule: Kone et al. (2025a) identify the Pareto set by posterior sampling in a transductive linear model, which handles structure and correlation between objectives at a small computational cost and is asymptotically optimal in both the frequentist and the Bayesian sense, and Kone et al. (2025b) extend identification to the constrained case, where some objectives act as feasibility thresholds rather than as criteria. A parallel line replaces componentwise dominance by an ordering cone, so that partial or incomplete preferences can be encoded in the relation itself (Ararat & Tekin, 2023; Karagözlü et al., 2024). These results are the reference point for what is achievable under noise, and they are complementary to ours: they assume a finite arm set and repeated pulls, and spend their budget on resolving gaps between known candidates, whereas MOBO must propose new designs in a continuous space from one evaluation each. The soft indicator

of Section 3.1 can be read as a cheap surrogate for the same membership question when that budget is unavailable.

## B NOISE STRUCTURES AND CALIBRATION

All expectations and ranges below are computed once per problem on a Sobol grid of 4096 points over the search space; $r_k$ is the range of $f_k$ on that grid. Each structure is set so that the noise factor (5) equals the requested $\eta$ for every objective.

**Constant.** $\varepsilon_k = \sigma_k z$ with $\sigma_k = \eta\, r_k$.

**Contaminated.** A Tukey–Huber mixture $\varepsilon_k \sim (1-p)\,\mathcal{N}(0,\sigma_k^2) + p\,\mathcal{N}(0,K^2\sigma_k^2)$ with $p = 0.1$ and $K = 5$, drawn independently per objective. Its variance is $\sigma_k^2(1-p+pK^2)$, so $\sigma_k = \eta\, r_k/\sqrt{1-p+pK^2}$. The variance does not depend on $\boldsymbol{x}$, but the marginal is heavy-tailed, so a Gaussian likelihood is misspecified even if its variance is correct.

**Multiplicative.** $y_k = f_k(\boldsymbol{x})\, g_k$ with $g_k = \exp(\sigma_k z - \sigma_k^2/2)$, so $\mathbb{E}[g_k] = 1$ and $\mathbb{V}[y_k \mid \boldsymbol{x}] = f_k(\boldsymbol{x})^2(e^{\sigma_k^2}-1)$. Averaging over $\boldsymbol{x}$ and solving (5) gives $\sigma_k^2 = \log(1+\eta^2 r_k^2/\mathbb{E}_{\boldsymbol{x}}[f_k(\boldsymbol{x})^2])$. The noise vanishes where $f_k = 0$ and grows with $|f_k|$.

**Hidden input.** The last input of the original problem $F\colon \mathbb{R}^d \to \mathbb{R}^m$ is withheld from the optimizer: at every evaluation it is drawn as $\xi \sim \mathrm{Beta}(5,5)$ rescaled to its bounds, and the optimizer controls only the remaining inputs $\boldsymbol{x}$. The target is $f_k(\boldsymbol{x}) = \mathbb{E}_\xi[F_k(\boldsymbol{x},\xi)]$, which we approximate by a quasi-Monte Carlo average over a fixed 64-point Sobol grid in $\xi$, and the observation is $y_k = f_k(\boldsymbol{x}) + a_k[F_k(\boldsymbol{x},\xi) - f_k(\boldsymbol{x})]$. The variance $a_k^2\,\mathbb{V}_\xi[F_k(\boldsymbol{x},\xi)]$ depends on $\boldsymbol{x}$ through the problem itself. The amplitude is $a_k = \eta\, r_k/\sqrt{\mathbb{E}_{\boldsymbol{x}}[\mathbb{V}_\xi[F_k(\boldsymbol{x},\xi)]]}$, where $r_k$ is the range of $f_k$ over the observed inputs and the expectation uses a 1024-point grid in $\xi$ to keep the calibration bias small.

## C THEORETICAL RESULTS

This appendix collects the formal statements used in the paper and their proofs.

### C.1 THE SOFT NON-DOMINANCE INDICATOR IN THE LIMIT

*Proof of Lemma 1.* Let $\mathcal{K}_i = \{\, j \le n : \boldsymbol{y}_j \succ \boldsymbol{y}_i \,\}$, so $k_i = |\mathcal{K}_i|$, and fix a subsample $I$ of size $s$ with front $P = \mathrm{ND}(\{\boldsymbol{y}_j\}_{j\in I})$.

We first show that some element of $P$ dominates $\boldsymbol{y}_i$ if and only if $I \cap \mathcal{K}_i \neq \varnothing$. If $j \in I \cap \mathcal{K}_i$, then either $\boldsymbol{y}_j \in P$, in which case $P$ contains a point dominating $\boldsymbol{y}_i$, or $\boldsymbol{y}_j$ is dominated by some $\boldsymbol{p} \in P$, since dominance is a strict partial order on a finite set; transitivity then gives $\boldsymbol{p} \succ \boldsymbol{y}_j \succ \boldsymbol{y}_i$. Conversely, an element of $P$ that dominates $\boldsymbol{y}_i$ is one of the $\boldsymbol{y}_j$ with $j \in I$ and $j \in \mathcal{K}_i$.

The event in (6) is therefore $I \cap \mathcal{K}_i = \varnothing$, which requires all $s$ draws to fall in the $n - k_i$ indices outside $\mathcal{K}_i$. Under uniform sampling without replacement its probability is the hypergeometric expression $\binom{n-k_i}{s}/\binom{n}{s}$, equal to the stated product, and zero when $n - k_i < s$. Note that $i \notin \mathcal{K}_i$, so whether $i$ itself is drawn does not matter. The average in (6) is a mean of $B$ independent Bernoulli draws of this event, so it converges to that probability as $B \to \infty$. □

### C.2 DOMINANCE PROPERTIES OF SUBSAMPLED ACQUISITIONS

**Population acquisition and scope.** We index acquisitions by the query they score. Before iteration $t$, the observed data are $\mathcal{D}_{t-1}$. Algorithm 1 draws index sets $\mathcal{I}^{(b)}$ from $\{1,\dots,t-1\}$ and forms the corresponding subsampled fronts $\mathcal{P}^{(b)} = \mathrm{ND}(\{\boldsymbol{y}_i : i \in \mathcal{I}^{(b)}\})$. For the analysis, it is convenient to retain the finite comparison set before taking its non-dominated subset. We therefore write

$$\mathcal{S}_{t-1}^{(b)} \coloneqq \{\boldsymbol{y}_i : i \in \mathcal{I}^{(b)}\}, \qquad \mathcal{P}^{(b)} = \mathrm{ND}(\mathcal{S}_{t-1}^{(b)}). \tag{18}$$

For the HVI-based utilities considered below, scoring against $\mathcal{S}_{t-1}^{(b)}$ is equivalent to scoring against $\mathcal{P}^{(b)}$, since dominated elements of the comparison set do not alter its dominated hypervolume.

Let $\mathcal{S}_{t-1}$ denote one random comparison set generated according to the same subsampling rule. The population acquisition corresponding to the Monte Carlo scores of Algorithm 1 is

$$a_t(\boldsymbol{x}) \coloneqq \mathbb{E}[u(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}(\boldsymbol{x}); \mathcal{S}_{t-1}) \mid \mathcal{D}_{t-1}]. \tag{19}$$

The expectation is over both the random comparison set and a fresh observation at $\boldsymbol{x}$. Averaging over the $B$ independent subsamples in Algorithm 1 gives a Monte Carlo approximation of this expectation. The fitted classifier acquisition $\widehat{a}_t$ further approximates this population quantity. Controlling that approximation, however, is outside the present analysis and reserved for future work.

**Standing assumptions.** Let $\mathcal{X}$ be a compact Polish[2] space equipped with its Borel $\sigma$-algebra. We assume there exists an underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ where the history $\mathfrak{H}_t$, the $\sigma$-algebra generated by the data $\mathcal{D}_t$, forms a filtration on $\mathfrak{A}$. We also assume that $\boldsymbol{f} : \mathcal{X} \to \mathcal{Y}$ is continuous on the domain $\mathcal{X}$, the codomain being $\mathcal{Y} = \mathbb{R}^m$, and that the observation noise has mean zero, with further assumptions presented as needed.

**Assumption 1** (Dominance-compatible utility)**.** *The utility $u : \mathcal{Y} \times 2^{\mathcal{Y}} \to [0, \infty)$ is jointly measurable. Namely, for any $n \in \mathbb{N}$, the mapping $(\boldsymbol{y}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \mapsto u(\boldsymbol{y}; \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\})$ is Borel measurable, for $\boldsymbol{y}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in \mathcal{Y}$. In addition, given any finite $\mathcal{S} \subset \mathcal{Y}$ and any $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}$,*

$$\boldsymbol{y}' \preceq \boldsymbol{y} \quad \Longrightarrow \quad u(\boldsymbol{y}'; \mathcal{S}) \le u(\boldsymbol{y}; \mathcal{S}), \tag{20}$$

$$\exists \boldsymbol{z} \in \mathcal{S} : \quad \boldsymbol{y} \preceq \boldsymbol{z} \quad \Longrightarrow \quad u(\boldsymbol{y}; \mathcal{S}) = 0, \tag{21}$$

$$\exists \alpha \ge 0 : \quad u(\boldsymbol{y}; \mathcal{S}) \le c_u(\boldsymbol{y})(1 + |\mathcal{S}|)^{\alpha}, \tag{22}$$

*where $c_u : \mathcal{Y} \to [0, \infty)$ is some finite measurable function.*

*Remark* 1. Assumption 1 ensures non-strict preservation of dominance under a common candidate-noise law. Letting $\boldsymbol{\varepsilon} \sim \rho_{\boldsymbol{\varepsilon}}$ have the same distribution for every $\boldsymbol{x} \in \mathcal{X}$, the monotonicity condition gives:

$$\boldsymbol{f}(\boldsymbol{x}') \preceq \boldsymbol{f}(\boldsymbol{x}) \quad \Longrightarrow \quad \int_{\mathcal{Y}} u(\boldsymbol{f}(\boldsymbol{x}') + \boldsymbol{z}; \mathcal{S}) \, \mathrm{d}\rho_{\boldsymbol{\varepsilon}}(\boldsymbol{z}) \le \int_{\mathcal{Y}} u(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{z}; \mathcal{S}) \, \mathrm{d}\rho_{\boldsymbol{\varepsilon}}(\boldsymbol{z}), \qquad \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}. \tag{23}$$

**Proposition 1.** *For fixed reference $\boldsymbol{r} \in \mathcal{Y}$, the utilities:*

$$u_{\mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}) \coloneqq \mathrm{HVI}(\boldsymbol{y}; \mathcal{S}), \tag{24}$$

$$u_{\mathrm{PHVI}}(\boldsymbol{y}; \mathcal{S}) \coloneqq \mathbb{1}[u_{\mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}) > 0], \tag{25}$$

$$u_{\log \mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}) \coloneqq \log(1 + u_{\mathrm{HVI}}(\boldsymbol{y}; \mathcal{S})) \tag{26}$$

*satisfy Assumption 1 with $\alpha = 0$.*

*Proof.* Given $\boldsymbol{y} \in \mathcal{Y}$, let $\mathcal{M}_{\boldsymbol{r}}(\boldsymbol{y}) \coloneqq [\boldsymbol{r}, \boldsymbol{y}]$ denote the region dominated by $\boldsymbol{y}$, and let $\mathcal{M}_{\boldsymbol{r}}(\mathcal{S}) \coloneqq \cup_{\boldsymbol{z} \in \mathcal{S}} \mathcal{M}_{\boldsymbol{r}}(\boldsymbol{z})$. If $\boldsymbol{y}' \preceq \boldsymbol{y}$, then $\mathcal{M}_{\boldsymbol{r}}(\boldsymbol{y}') \subseteq \mathcal{M}_{\boldsymbol{r}}(\boldsymbol{y})$, and hence:

$$u_{\mathrm{HVI}}(\boldsymbol{y}'; \mathcal{S}) = \int_{\mathcal{M}_{\boldsymbol{r}}(\boldsymbol{y}') \setminus \mathcal{M}_{\boldsymbol{r}}(\mathcal{S})} \mathrm{d}\boldsymbol{z} \le \int_{\mathcal{M}_{\boldsymbol{r}}(\boldsymbol{y}) \setminus \mathcal{M}_{\boldsymbol{r}}(\mathcal{S})} \mathrm{d}\boldsymbol{z} = u_{\mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}). \tag{27}$$

If $\boldsymbol{y} \preceq \boldsymbol{y}'$ for some $\boldsymbol{y}' \in \mathcal{S}$, then $\mathcal{M}_{\boldsymbol{r}}(\boldsymbol{y}) \subseteq \mathcal{M}_{\boldsymbol{r}}(\mathcal{S})$ and $u_{\mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}) = 0$. Therefore, the first two properties in Assumption 1 hold for $u_{\mathrm{HVI}}$, and transfer directly to $u_{\mathrm{PHVI}}$ and $u_{\log \mathrm{HVI}}$. Furthermore,

$$u_{\mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}) \le \mathrm{HV}(\{\boldsymbol{y}\}) = \prod_{i=1}^{m} \max(\boldsymbol{y}_i - \boldsymbol{r}_i, 0), \tag{28}$$

$$u_{\mathrm{PHVI}}(\boldsymbol{y}; \mathcal{S}) \le 1, \tag{29}$$

$$u_{\log \mathrm{HVI}}(\boldsymbol{y}; \mathcal{S}) \le \log\left(1 + \prod_{i=1}^{m} \max(\boldsymbol{y}_i - \boldsymbol{r}_i, 0)\right), \tag{30}$$

which satisfy the envelope condition in Assumption 1. Lastly, $u_{\mathrm{HVI}}$ is Borel measurable in the candidate and the elements of any finite comparison set. The other two utilities are Borel measurable as compositions of $u_{\mathrm{HVI}}$ with Borel-measurable functions. □

[2] A Polish space is a separable completely metrizable topological space (i.e., its topology can be induced by a complete metric), which includes $\mathbb{R}^d$, compact metric spaces, and finite or countable discrete sets.

The next result shows that Lemma 1, originally formulated for indicator-based HVI utilities, extends to the more general class of utilities considered here. In particular, the same decay with respect to the number of observed dominators applies to LF-NEHVI and similar utilities, becoming geometric with a fixed rate under proportional subsampling.

**Lemma 2.** *Let $u : \mathcal{Y} \times 2^{\mathcal{Y}} \to [0, \infty)$ satisfy Assumption 1. Conditional on $\mathfrak{H}_t$, let $\mathcal{S}_t$ be obtained by uniformly sampling $n_t$ observation indices without replacement from $\mathcal{Y}_t$, where $0 \le n_t \le t$ may depend on $t$. For $t \ge 1$ and $\boldsymbol{y} \in \mathcal{Y}$, let:*

$$U_t(\boldsymbol{y}) \coloneqq \max_{\mathcal{I} \subseteq \{1,\dots,t\}:\, |\mathcal{I}| = n_t} u(\boldsymbol{y}; \{\boldsymbol{y}_i : i \in \mathcal{I}\}), \qquad K_t(\boldsymbol{y}) \coloneqq \sum_{i=1}^{t} \mathbb{1}[\boldsymbol{y} \prec \boldsymbol{y}_i]. \tag{31}$$

*Then, almost surely, for each fixed $\boldsymbol{y} \in \mathcal{Y}$,*

$$\mathbb{E}[u(\boldsymbol{y}; \mathcal{S}_t) \mid \mathfrak{H}_t] \le U_t(\boldsymbol{y}) \Big(1 - \frac{n_t}{t}\Big)^{K_t(\boldsymbol{y})}, \tag{32}$$

*with $0^0 \coloneqq 1$ by convention.*

*Proof.* Fix $t \ge 1$ and $\boldsymbol{y} \in \mathcal{Y}$, and let $\mathcal{I}_t^\star(\boldsymbol{y}) \coloneqq \{i \le t : \boldsymbol{y} \prec \boldsymbol{y}_i\}$. For every $\mathcal{I} \subseteq \{1, \dots, t\}$ of size $n_t$, the zeroing property gives:

$$u(\boldsymbol{y}; \{\boldsymbol{y}_i : i \in \mathcal{I}\}) \le U_t(\boldsymbol{y}) \mathbb{1}[\mathcal{I} \cap \mathcal{I}_t^\star(\boldsymbol{y}) = \varnothing]. \tag{33}$$

Let $\widetilde{\mathcal{I}}_t$ denote the uniformly sampled index set defining $\mathcal{S}_t$. Conditional on $\mathfrak{H}_t$, we have $|\mathcal{I}_t^\star(\boldsymbol{y})| = K_t(\boldsymbol{y})$, and hence:

$$\begin{aligned} \mathbb{E}[u(\boldsymbol{y}; \mathcal{S}_t) \mid \mathfrak{H}_t] &\le U_t(\boldsymbol{y}) \mathbb{P}\Big[\widetilde{\mathcal{I}}_t \cap \mathcal{I}_t^\star(\boldsymbol{y}) = \varnothing \mid \mathfrak{H}_t\Big] \\ &= U_t(\boldsymbol{y}) \frac{\binom{t - K_t(\boldsymbol{y})}{n_t}}{\binom{t}{n_t}} \le U_t(\boldsymbol{y}) \Big(1 - \frac{n_t}{t}\Big)^{K_t(\boldsymbol{y})}, \end{aligned} \tag{34}$$

which concludes the proof. □

As Lemma 2 shows, observed domination suppresses the subsampled utility geometrically with the number of observed dominators when $n_t = \lfloor \gamma t \rfloor$ for some $\gamma \in (0, 1)$. The next result establishes a complementary property: under a common full-support candidate-noise law $\boldsymbol{\varepsilon} \sim \rho_{\boldsymbol{\varepsilon}}$, HVI-based expected utilities strictly preserve dominance in the underlying objectives. Unlike the preceding result, it does not depend on how the comparison set was sampled.

**Lemma 3** (Strict expected dominance)**.** *Fix a reference point $\boldsymbol{r} \in \mathcal{Y}$ and designs $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ satisfying $\boldsymbol{f}(\boldsymbol{x}') \preceq \boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x}') \ne \boldsymbol{f}(\boldsymbol{x})$. Suppose $m \ge 2$, $\rho_{\boldsymbol{\varepsilon}}$ is a full-support probability measure on $\mathcal{Y}$ common to all candidate designs, and $u$ satisfies Assumption 1. Assume that:*

$$u(\boldsymbol{y}; \mathcal{S}) = \psi(\mathrm{HVI}(\boldsymbol{y}; \mathcal{S})), \tag{35}$$

*for every $\boldsymbol{y} \in \mathcal{Y}$ and finite $\mathcal{S} \subset \mathcal{Y}$, where $\psi : [0, \infty) \to [0, \infty)$ is nondecreasing, $\psi(0) = 0$, and $\psi(s) > 0$ for every $s > 0$. If the relevant expected utilities are finite, then, for every finite $\mathcal{S} \subset \mathcal{Y}$,*

$$\int_{\mathcal{Y}} u(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{z}; \mathcal{S}) \, \mathrm{d}\rho_{\boldsymbol{\varepsilon}}(\boldsymbol{z}) > \int_{\mathcal{Y}} u(\boldsymbol{f}(\boldsymbol{x}') + \boldsymbol{z}; \mathcal{S}) \, \mathrm{d}\rho_{\boldsymbol{\varepsilon}}(\boldsymbol{z}). \tag{36}$$

*Proof.* Let $\boldsymbol{d} \coloneqq \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}')$. Choose $\ell \in \{1, \dots, m\}$ such that $d_\ell > 0$, and choose $\delta > 0$ such that $2\delta < d_\ell$. Since $\mathcal{S}$ is finite, for each $j \ne \ell$ choose $q_j$ such that:

$$q_j > \max(\{r_j\} \cup \{y_j : \boldsymbol{y} \in \mathcal{S}\}) + 1. \tag{37}$$

Define:

$$\mathcal{Q} \coloneqq [r_\ell, r_\ell + \delta] \times \prod_{j \ne \ell} [q_j, q_j + \delta], \qquad \mathcal{A} \coloneqq (r_\ell - d_\ell + \delta, r_\ell - \delta) \times \prod_{j \ne \ell} (q_j + \delta, q_j + 2\delta). \tag{38}$$

The set $\mathcal{E} \coloneqq \mathcal{A} - \boldsymbol{f}(\boldsymbol{x}')$ is nonempty and open, so full support gives $\rho_{\boldsymbol{\varepsilon}}(\mathcal{E}) > 0$. For $\boldsymbol{z} \in \mathcal{E}$, set $\boldsymbol{y}' \coloneqq \boldsymbol{f}(\boldsymbol{x}') + \boldsymbol{z}$ and $\boldsymbol{y} \coloneqq \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{z} = \boldsymbol{y}' + \boldsymbol{d}$. Then $y'_\ell < r_\ell$, implying $\mathrm{HVI}(\boldsymbol{y}'; \mathcal{S}) = 0$. Also,

$\boldsymbol{y}$ dominates $\mathcal{Q}$, while no element of $\mathcal{S}$ dominates any point of $\mathcal{Q}$ by the choice of $q_j$ for $j \neq \ell$. Consequently,

$$\mathrm{HVI}(\boldsymbol{y}; \mathcal{S}) \geq \mathrm{vol}(\mathcal{Q}) = \delta^m, \tag{39}$$

and hence:

$$u(\boldsymbol{y}; \mathcal{S}) - u(\boldsymbol{y}'; \mathcal{S}) \geq \psi(\delta^m) > 0. \tag{40}$$

For $\boldsymbol{z} \notin \mathcal{E}$, the same utility difference is nonnegative by the dominance monotonicity in Assumption 1. Therefore,

$$\int_{\mathcal{Y}} (u(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{z}; \mathcal{S}) - u(\boldsymbol{f}(\boldsymbol{x}') + \boldsymbol{z}; \mathcal{S}))\, \mathrm{d}\rho_{\boldsymbol{\varepsilon}}(\boldsymbol{z}) \geq \rho_{\boldsymbol{\varepsilon}}(\mathcal{E})\psi(\delta^m) > 0, \tag{41}$$

which concludes the proof. □

**Candidate-independent subsampling.** The following Pareto-alignment result does not require the particular uniform without-replacement subsampling scheme used in Algorithm 1. We call a subsampling scheme *candidate-independent* if, conditional on $\mathfrak{H}_{t-1}$, the distribution of the random finite comparison set $\mathcal{S}_{t-1} \subseteq \mathcal{Y}_{t-1}$ does not depend on the candidate $\boldsymbol{x}$ being scored. The scheme is therefore allowed to depend on the observed history, and its subsample size may vary with $t$. Uniform sampling without replacement with $|\mathcal{S}_{t-1}| = \lfloor \gamma(t-1) \rfloor$ we used is one practical special case.

**Proposition 2** (Pareto alignment of the population acquisition)**.** *Suppose the conditions of Lemma 3 hold for a given utility $u$, objective $\boldsymbol{f}$, and common candidate-noise law $\rho_{\boldsymbol{\varepsilon}}$. For each $t \in \mathbb{N}$, let $\mathcal{S}_{t-1}$ be generated by any candidate-independent subsampling scheme, and let the fresh observation noise $\boldsymbol{\varepsilon} \sim \rho_{\boldsymbol{\varepsilon}}$ be conditionally independent of $\mathcal{S}_{t-1}$ given $\mathfrak{H}_{t-1}$. Assume the corresponding conditional expected utilities are finite. Then, almost surely, for every $t \in \mathbb{N}$,*

$$\boldsymbol{f}(\boldsymbol{x}') \preceq \boldsymbol{f}(\boldsymbol{x}) \ \wedge \ \boldsymbol{f}(\boldsymbol{x}') \neq \boldsymbol{f}(\boldsymbol{x}) \quad \Longrightarrow \quad a_t(\boldsymbol{x}') < a_t(\boldsymbol{x}), \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}. \tag{42}$$

*Consequently, every maximizer of $a_t$ is Pareto optimal:*

$$\arg\max_{\boldsymbol{x} \in \mathcal{X}} a_t(\boldsymbol{x}) \subseteq \mathcal{P}^\star, \qquad t \in \mathbb{N}. \tag{43}$$

*Proof.* Fix $t \in \mathbb{N}$ and $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ such that $\boldsymbol{f}(\boldsymbol{x}') \preceq \boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x}') \neq \boldsymbol{f}(\boldsymbol{x})$. Conditional on $\mathfrak{H}_{t-1}$, candidate independence ensures that the same distribution over $\mathcal{S}_{t-1}$ is used to score both designs. By conditional independence of the fresh noise and Lemma 3,

$$a_t(\boldsymbol{x}) - a_t(\boldsymbol{x}') = \mathbb{E}\left[ \int_{\mathcal{Y}} \Big( u(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{z}; \mathcal{S}_{t-1}) - u(\boldsymbol{f}(\boldsymbol{x}') + \boldsymbol{z}; \mathcal{S}_{t-1}) \Big)\, \mathrm{d}\rho_{\boldsymbol{\varepsilon}}(\boldsymbol{z}) \,\middle|\, \mathfrak{H}_{t-1} \right] > 0, \tag{44}$$

since the inner integral is strictly positive for every finite realization of $\mathcal{S}_{t-1}$ by Lemma 3. Thus $a_t$ strictly preserves every strict Pareto-dominance relation.

To prove the final inclusion, let $\boldsymbol{x}' \in \arg\max_{\boldsymbol{x} \in \mathcal{X}} a_t(\boldsymbol{x})$. Suppose, for contradiction, that $\boldsymbol{x}' \notin \mathcal{P}^\star$. By the definition of the Pareto set, there then exists some $\boldsymbol{x} \in \mathcal{X}$ such that $\boldsymbol{f}(\boldsymbol{x}') \preceq \boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x}') \neq \boldsymbol{f}(\boldsymbol{x})$. The strict dominance preservation established above implies $a_t(\boldsymbol{x}') < a_t(\boldsymbol{x})$, contradicting the maximality of $\boldsymbol{x}'$. Hence $\boldsymbol{x}' \in \mathcal{P}^\star$, and therefore,

$$\arg\max_{\boldsymbol{x} \in \mathcal{X}} a_t(\boldsymbol{x}) \subseteq \mathcal{P}^\star. \tag{45}$$

which concludes the proof. □

### C.3 Ordering of the evaluation conventions

**Proposition 3** (Ordering and decomposition)**.** *For every run, checkpoint $n$ and estimator $E$, $\mathrm{IGD}^{+,\mathrm{orc}}_n \leq \mathrm{IGD}^{+,E}_n \leq \mathrm{nIGD}^{+,E}_n$ and $\mathrm{HV}^{E}_n \leq \mathrm{HV}^{\mathrm{orc}}_n$. If $\mathcal{P}_n \subseteq \widehat{\mathcal{S}}^{E}_n$, then $\mathrm{IGD}^{+,E}_n = \mathrm{IGD}^{+,\mathrm{orc}}_n$ and $\mathrm{HV}^{E}_n = \mathrm{HV}^{\mathrm{orc}}_n$. Consequently the decomposition (17) holds with all three of its terms nonnegative.*

**Algorithm 2** Sequential LF-MOBO

**Require:** Initial dataset $\mathcal{D}_0$, evaluation budget $T$, subsample fraction $\gamma$, number of subsamples $B$, reference point $\boldsymbol{r}$ for LF-NEHVI

1: **for** $n = 0, \dots, T-1$ **do**
2: Run Algorithm 1 on $\mathcal{D}_n$ to select $\boldsymbol{x}_{n+1}$
3: Query $\boldsymbol{x}_{n+1}$ once and observe $\boldsymbol{y}_{n+1} = \boldsymbol{f}(\boldsymbol{x}_{n+1}) + \boldsymbol{\varepsilon}(\boldsymbol{x}_{n+1})$
4: $\mathcal{D}_{n+1} \leftarrow \mathcal{D}_n \cup \{(\boldsymbol{x}_{n+1}, \boldsymbol{y}_{n+1})\}$
5: **end for**
6: **return** $\mathcal{D}_T$

*Proof.* The first inequality holds because the minimum in the returned-set IGD$^+$ runs over $\widehat{\mathcal{S}}_n^E \subseteq \{1, \dots, n\}$, a subset of the indices in (11). The second holds term by term: $j^E(\boldsymbol{a}) \in \widehat{\mathcal{S}}_n^E$, so $d^+(\boldsymbol{a}, \boldsymbol{f}(\boldsymbol{x}_{j^E(\boldsymbol{a})})) \geq \min_{i \in \widehat{\mathcal{S}}_n^E} d^+(\boldsymbol{a}, \boldsymbol{f}(\boldsymbol{x}_i))$, whatever rule selects $j^E$. Hypervolume is monotone under set inclusion, which gives $\mathrm{HV}_n^E \leq \mathrm{HV}_n^{\mathrm{orc}}$. Equation (17) is then the identity $\mathrm{nIGD}^+ = \mathrm{IGD}^{+,\mathrm{orc}} + (\mathrm{IGD}^{+,E} - \mathrm{IGD}^{+,\mathrm{orc}}) + (\mathrm{nIGD}^{+,E} - \mathrm{IGD}^{+,E})$ with non-negative brackets.

For the equality case, suppose $\mathcal{P}_n \subseteq \widehat{\mathcal{S}}_n^E$. Since dominance is a strict partial order on a finite set, every evaluated $\boldsymbol{f}(\boldsymbol{x}_i)$ is weakly dominated by $\boldsymbol{f}(\boldsymbol{x}_p)$ for some $p \in \mathcal{P}_n$. Then $[\boldsymbol{r}, \boldsymbol{f}(\boldsymbol{x}_i)] \subseteq [\boldsymbol{r}, \boldsymbol{f}(\boldsymbol{x}_p)]$, so dropping the designs outside $\widehat{\mathcal{S}}_n^E$ leaves the union of boxes, and hence the hypervolume, unchanged; and $d^+(\boldsymbol{a}, \boldsymbol{f}(\boldsymbol{x}_p)) \leq d^+(\boldsymbol{a}, \boldsymbol{f}(\boldsymbol{x}_i))$ for every target $\boldsymbol{a}$, because $d^+$ is non-increasing in each objective, so the minimum over $i \leq n$ is already attained in $\mathcal{P}_n \subseteq \widehat{\mathcal{S}}_n^E$. □

## D Implementation details

**Algorithm.** Algorithm 2 describes the outer BO loop and how it connects with the candidate-selection step of Algorithm 1. Starting from the initial dataset $\mathcal{D}_0$, each round $n$ calls Algorithm 1 to propose the next candidate $\boldsymbol{x}_{n+1}$, which is done by computing utility weights from the current observations in $\mathcal{D}_n$, fitting the classifier, and maximizing over its predictions. The black-box objective $\boldsymbol{f}$ is then evaluated once at $\boldsymbol{x}_{n+1}$, and the resulting observation $\boldsymbol{y}_{n+1}$ is added to $\mathcal{D}_n$ before the next round. Classifier training and acquisition maximization occur within each round, proposing and evaluating one candidate per round.

**Runs.** Every problem is cast as maximization. A run draws $n_0 = 2(d+1)$ Sobol designs and then evaluates 100 designs one at a time ($q = 1$), so a trial ends with $n_0 + 100$ evaluations, each observed once. A single seed fixes the initial design, the noise draws, the Monte Carlo samplers and the subsampling of Algorithm 1; we use ten seeds per condition. Problems are the BoTorch implementations (for WFG6, the pymoo implementation (Blank & Deb, 2020)), with the reference point of Daulton et al. (2021) where the two differ, and the reference fronts are the analytic fronts when available and otherwise the non-dominated subset of a $2^{16}$-point Sobol grid, subsampled to at most 5000 points.

**GP baselines.** Each objective gets a `SingleTaskGP` with standardized outputs on inputs mapped to the unit cube, collected in a `ModelListGP` and fitted by maximum a posteriori estimation with BoTorch's default priors (Balandat et al., 2020); no method is given the noise variances, so the noise level is inferred along with the other hyperparameters. The acquisitions are qLogEHVI on a box decomposition of the observed front, qLogNEHVI and qLogNParEGO with the evaluated designs as baseline and baseline pruning, all with 128 quasi-Monte Carlo samples, and the BoTorch implementation of qLB-JES (Tu et al., 2022), used without tuning, with 8 sampled Pareto sets of 8 points each, obtained by random search, and the lower-bound estimator. Acquisitions are maximized over the unit cube with `optimize_acqf`, 10 restarts and 256 raw samples, which is also how the likelihood-free acquisitions are maximized.

**Likelihood-free methods.** The weights of Algorithm 1 use $\gamma = 1/2$ and $B = 128$ subsamples, redrawn every round. The classifier is an ensemble of five multilayer perceptrons, each three pre-norm residual blocks wide $\min(256, \max(32, 8d))$ with SiLU activations, dropout 0.2 and an inner

expansion factor of two, and each regularized toward its own random initialization following Pearce et al. (2020), with the regularization coefficient schedule $\lambda_t = \lambda_0(\log(t+2))^q$ at $\lambda_0 = 0.25$ (for iteration, $t$) following the analysis in Oliveira (2026). Inputs are rescaled to $[-1, 1]$ to properly interact with the non-linear SiLU activations. Training minimizes (3) for 50 epochs with AdamW at learning rate $10^{-3}$, mini-batches of at most 32, gradient clipping at norm 1, and the weights rescaled to mean one; the output bias starts at the pointwise optimum of the rescaled weights. The acquisition is the mean logit of the ensemble. Nothing in this configuration is tuned per problem.

**Evaluation.** Oracle and likelihood-free metrics are computed at every iteration and the VHGP metrics every ten, refitting the estimator on the data available at that iteration. The VHGP fits follow Lázaro-Gredilla & Titsias (2011): inputs rescaled by the lengthscales of a homoscedastic ARD fit, then L-BFGS on the variational parameters for 60 iterations and jointly with the hyperparameters for 120, with 1024 posterior samples for the membership probabilities. Hypervolume is computed exactly with `moocore`, which is correct for four or more objectives on point sets with repeated coordinates (Appendix E). Every VHGP fit completed, and 97.9% of them converged. Four qLogNEHVI cells on the $m = 5$ problems have 7 to 9 seeds rather than 10.

## E HYPERVOLUME COMPUTATION

For computing the hypervolume performance metrics post-run we use the exact computation provided by `moocore` (https://github.com/multi-objective/moocore), which implements the dimension-sweep algorithms of Fonseca et al. (2006) and Guerreiro & Fonseca (2018). We measured it against the exact `Hypervolume` of BoTorch (Balandat et al., 2020). Table 3 reports the median time per call on one CPU core.

The speedup grows with both the number of objectives and the size of the front, from about $10^2$ at $m = 2$ to more than $4 \times 10^4$ at $m = 5$ with a 100-point front, because the two implementations scale differently in $m$. Scoring this campaign takes about 2.9 million hypervolume evaluations: the oracle and likelihood-free metrics at every one of the 101 iterations of each of the 13,432 runs, and the VHGP metrics at every tenth. At the median front size of each problem, that is about half a minute of hypervolume computation in total with `moocore`, against an estimated 4.4 CPU-days with BoTorch, which is what makes the per-iteration protocol of Section 4 affordable at all.

Table 3: Median time of one exact hypervolume evaluation on a front of $n$ mutually non-dominated points in $m$ objectives, and the speedup of `moocore` over the implementation in BoTorch; the faster time of each row is in bold. Both compute the same quantity on the same data, on one CPU core (x86_64) with BoTorch 0.18.1, `moocore` 0.3.2 and Python 3.11.15.

| $m$ | $n$ | `moocore` (ms) | BoTorch (ms) | speedup |
|---|---|---|---|---|
| 2 | 25 | **0.007** | 0.7 | 105 |
| | 50 | **0.007** | 1.6 | 219 |
| | 100 | **0.009** | 3.6 | 406 |
| 3 | 25 | **0.009** | 5.0 | 556 |
| | 50 | **0.010** | 14.7 | 1404 |
| | 100 | **0.014** | 51.4 | 3648 |
| 4 | 25 | **0.008** | 18.5 | 2339 |
| | 50 | **0.012** | 88.3 | 7543 |
| | 100 | **0.015** | 529.2 | 34770 |
| 5 | 25 | **0.012** | 138.4 | 11510 |
| | 50 | **0.023** | 918.9 | 40348 |
| | 100 | **0.164** | 7265.6 | 44373 |

We retain the original BoTorch hypervolume computations (which allow for approximations in high-$m$) for runs of all EHVI-based methods, including our likelihood free.

## F COST OF A ROUND AND ABLATIONS

**Cost of a round.** The weights of LF-NPHVI cost $O(Bn^2m)$ dominance tests, or $O(n^2m)$ through Lemma 1, and involve no hypervolume computation. LF-NEHVI additionally needs one box decomposition of $P^{(b)}$ per subsample and one hypervolume evaluation per front member, which grows quickly with $m$ in the worst case, so we use it for moderate numbers of objectives ($m \leq 5$ here). Both replace the $m$ GP fits of a round, and their cost does not depend on how the noise is distributed. Table 4 reports the measured wall-clock time per iteration.

Table 4: Median wall-clock seconds per BO iteration, by number of objectives, over all runs of the campaign; the fastest method of each column is in bold. Sobol is below 0.05 seconds everywhere.

| | $m = 2$ | $m = 3$ | $m = 4$ | $m = 5$ |
|---|---|---|---|---|
| qLogEHVI | **3.3** | **5.1** | 23.9 | 55.5 |
| qLogNEHVI | 4.5 | 10.0 | 60.1 | 192.9 |
| qLogNParEGO | 4.5 | 5.9 | 5.3 | 8.1 |
| qLB-JES | 8.1 | 10.2 | 13.8 | 28.3 |
| LF-NPHVI | 7.8 | 5.7 | **3.3** | **4.3** |
| LF-NEHVI | 7.2 | 7.6 | 46.0 | 29.2 |

**Setup.** We vary $B$ and $\gamma$ one at a time around the defaults of Algorithm 1, on DTLZ2 ($d = 15$, $m = 3$), ZDT2 ($d = 20$, $m = 2$) and WFG6 ($d = 16$, $m = 4$), under constant noise at $\eta = 0.05$, contaminated noise at $\eta = 0.4$, and for DTLZ2 also hidden inputs at $\eta \in \{0.2, 0.4\}$, with the ten seeds of Section 5: 1,080 runs. The seeds are common random numbers across the settings, so the tests below are paired Wilcoxon signed-rank tests at the 5% level. We score the ablation with the HV gap and IGD$^+$ only. Both parameters act on the search, and those two metrics measure what the search finds; nIGD$^+$ adds the loss the decision maker incurs when picking one design out of the returned set, which is a property of the evaluation convention rather than of $B$ and $\gamma$, and is the subject of Section 5.

**Number of subsamples.** $B$ is a Monte Carlo budget rather than a modeling choice: by Lemma 1 the indicator has a closed-form $B \to \infty$ limit, so $B$ governs only the sampling error around a quantity that the data already determine, and that error is at most $1/(2\sqrt{B})$, about 0.04 at $B = 128$. Table 5 agrees: no comparison between $B = 1$ and the largest $B$ is significant, for either method or either metric (smallest $p = 0.07$), and the entries are not monotone in $B$. Cost does grow with $B$, but only for LF-NEHVI, which decomposes one box per subsample. We keep $B = 128$, which holds the sampling error of the labels below 0.05 at a cost that stays moderate for $m \leq 5$; LF-NPHVI can avoid the sampling altogether by evaluating (7) directly.

Table 5: Sensitivity of the likelihood-free methods to the number of subsamples $B$ and the subsample fraction $\gamma$, each varied around the default $B = 128$, $\gamma = 1/2$; the $\gamma = 1/2$ column is therefore the $B = 128$ arm and not a separate campaign. Quality entries are means over ten seeds and over the noise cells of the ablation (constant $\eta = 0.05$ and contaminated $\eta = 0.4$, and for DTLZ2 also hidden inputs at $\eta = 0.2$ and $0.4$); lower is better throughout, and the best entry of each row within each sweep is in bold. A run is 100 BO iterations. $B = 512$ was not run on WFG6, where a run at the default already takes hours.

| | | subsamples $B$ (at $\gamma = 1/2$) | | | | | fraction $\gamma$ (at $B = 128$) | | |
|---|---|---|---|---|---|---|---|---|---|
| problem | method | 1 | 8 | 32 | 128 | 512 | 0.25 | 0.5 | 0.75 |
| $\log_{10}$ *HV gap* | | | | | | | | | |
| DTLZ2 $d = 15$, $m = 3$ | LF-NPHVI | -0.162 | -0.168 | **-0.171** | -0.170 | -0.168 | -0.168 | -0.170 | **-0.171** |
| | LF-NEHVI | -0.154 | -0.156 | **-0.158** | -0.157 | -0.156 | -0.154 | **-0.157** | -0.155 |
| ZDT2 $d = 20$, $m = 2$ | LF-NPHVI | **1.193** | 1.270 | 1.235 | 1.262 | 1.252 | **1.195** | 1.262 | 1.309 |
| | LF-NEHVI | 1.246 | 1.310 | 1.247 | 1.187 | **1.154** | **1.047** | 1.187 | 1.265 |
| WFG6 $d = 16$, $m = 4$ | LF-NPHVI | 2.225 | 2.220 | **2.216** | 2.219 | – | **2.196** | 2.219 | 2.227 |
| | LF-NEHVI | 2.234 | 2.227 | 2.229 | **2.218** | – | 2.233 | **2.218** | 2.235 |
| *IGD*$^+$ | | | | | | | | | |
| DTLZ2 $d = 15$, $m = 3$ | LF-NPHVI | 0.492 | 0.474 | **0.456** | 0.469 | 0.471 | 0.466 | 0.469 | **0.461** |
| | LF-NEHVI | 0.539 | 0.533 | **0.531** | 0.532 | 0.539 | 0.535 | **0.532** | 0.538 |
| ZDT2 $d = 20$, $m = 2$ | LF-NPHVI | 1.293 | **1.224** | 1.387 | 1.408 | 1.330 | **1.169** | 1.408 | 1.413 |
| | LF-NEHVI | 1.609 | 1.605 | 1.500 | 1.461 | **1.424** | **1.210** | 1.461 | 1.618 |
| WFG6 $d = 16$, $m = 4$ | LF-NPHVI | 0.059 | 0.058 | **0.056** | 0.057 | – | **0.053** | 0.057 | 0.060 |
| | LF-NEHVI | 0.060 | 0.057 | 0.057 | **0.055** | – | 0.058 | **0.055** | 0.056 |
| *median wall-clock seconds of a run* | | | | | | | | | |
| DTLZ2 $d = 15$, $m = 3$ | LF-NPHVI | 591 | 436 | 441 | **433** | 459 | 487 | **433** | 474 |
| | LF-NEHVI | 440 | **430** | 444 | 502 | 676 | 534 | **502** | 585 |
| ZDT2 $d = 20$, $m = 2$ | LF-NPHVI | 945 | **939** | 1 095 | 1 090 | 1 106 | **762** | 1 090 | 826 |
| | LF-NEHVI | **1 028** | 1 085 | 1 093 | 1 102 | 1 106 | **970** | 1 102 | 1 056 |
| WFG6 $d = 16$, $m = 4$ | LF-NPHVI | **653** | 703 | 734 | 728 | – | **536** | 728 | 615 |
| | LF-NEHVI | **735** | 1 226 | 3 004 | 12 174 | – | **3 660** | 12 174 | 27 039 |

**Subsample fraction.** Unlike $B$, $\gamma$ enters the estimand itself, through $s = \lfloor \gamma n \rfloor$ in (7); we use the half-sample $\gamma = 1/2$ of Section 3.1 and do not tune it per problem. Terminal quality varies little over the range we tried, and $\gamma = 1/2$ is not the worst of the three settings for any combination of problem, method and metric. Four comparisons between $\gamma = 0.25$ and $\gamma = 0.75$ are significant, on WFG6 for LF-NPHVI ($p = 0.001$ and $p = 0.0006$) and on ZDT2 for LF-NEHVI ($p = 0.0006$ and $p = 0.0003$) and LF-NPHVI ($p = 0.03$ on IGD$^+$), in each case with the smaller fraction ahead; on DTLZ2 none is. The cost of $\gamma$, as of $B$, falls on LF-NEHVI at $m = 4$.

# G Additional results

## G.1 Wins, ties and losses against qLogNEHVI

Table 6 compares each likelihood-free method with qLogNEHVI condition by condition, by input-dimension group and noise level, under each of the three estimators.

Table 6: Wins/ties/losses of the likelihood-free methods against qLogNEHVI over conditions, by input-dimension group and noise level, under all three estimators. A win requires a two-sided Mann–Whitney $U$ test over seeds at the $5\%$ level. In each column and block, the likelihood-free method with the larger number of wins minus losses is in bold. Per noise level the groups hold 12 (low, $d \leq 7$) and 52 (high, $d \geq 8$; 48 for HV gap, where DTLZ7 $m$=5 is excluded) conditions. nIGD$^+$ needs an estimate of the objectives, so the oracle block reports IGD$^+$ in its place.

| | | $\eta = 0.05$ | | $\eta = 0.2$ | | $\eta = 0.4$ | |
|---|---|---|---|---|---|---|---|
| | Dimension, $d \rightarrow$ | low | high | low | high | low | high |
| LF est. | $\log_{10}$ HV gap | | | | | | |
| | LF-NPHVI | **0/4/8** | **5/25/18** | **1/8/3** | **29/15/4** | **4/7/1** | 26/18/4 |
| | LF-NEHVI | **0/4/8** | 1/18/29 | 0/8/4 | 27/17/4 | 1/9/2 | **29/16/3** |
| | nIGD$^+$ | | | | | | |
| | LF-NPHVI | **0/1/11** | **6/31/15** | **1/8/3** | **19/27/6** | **3/8/1** | 18/23/11 |
| | LF-NEHVI | **0/1/11** | 1/26/25 | 1/7/4 | 10/38/4 | 2/9/1 | **19/26/7** |
| VHGP est. | $\log_{10}$ HV gap | | | | | | |
| | LF-NPHVI | **0/4/8** | **7/24/17** | **0/7/5** | **24/20/4** | **4/5/3** | **21/24/3** |
| | LF-NEHVI | **0/4/8** | 2/19/27 | 0/6/6 | 18/27/3 | 1/9/2 | 21/23/4 |
| | nIGD$^+$ | | | | | | |
| | LF-NPHVI | **0/4/8** | 0/29/23 | **0/8/4** | **14/31/7** | 1/11/0 | 12/29/11 |
| | LF-NEHVI | **0/4/8** | **2/26/24** | 0/7/5 | 11/35/6 | **2/10/0** | **14/27/11** |
| oracle | $\log_{10}$ HV gap | | | | | | |
| | LF-NPHVI | **0/4/8** | **4/26/18** | **0/7/5** | **34/10/4** | **3/7/2** | 29/16/3 |
| | LF-NEHVI | **0/4/8** | 1/18/29 | **0/7/5** | 27/16/5 | 1/9/2 | **33/13/2** |
| | IGD$^+$ | | | | | | |
| | LF-NPHVI | **0/0/12** | **1/31/20** | **1/2/9** | **23/24/5** | **3/3/6** | 28/15/9 |
| | LF-NEHVI | **0/0/12** | 0/18/34 | 0/1/11 | 14/31/7 | 0/5/7 | **27/22/3** |

### G.2 Convergence curves

Figures 2 to 4 show convergence under each of the three estimators. The three figures carry the same conditions in the same grid positions, so a panel can be read across them and only the Pareto-set estimator changes; the metric is constant along a row, with two rows per metric. Every panel is a different condition, drawn as the mean over ten seeds with a $\pm 2$ standard error band; the VHGP values are computed every ten iterations. The conditions were chosen among those where neither likelihood-free method finishes significantly behind the best baseline under any estimator, for variety across problems, noise structures, noise levels and input dimension, which is why the curves are not all equally smooth. The terminal values of every condition are in Appendix G.6.

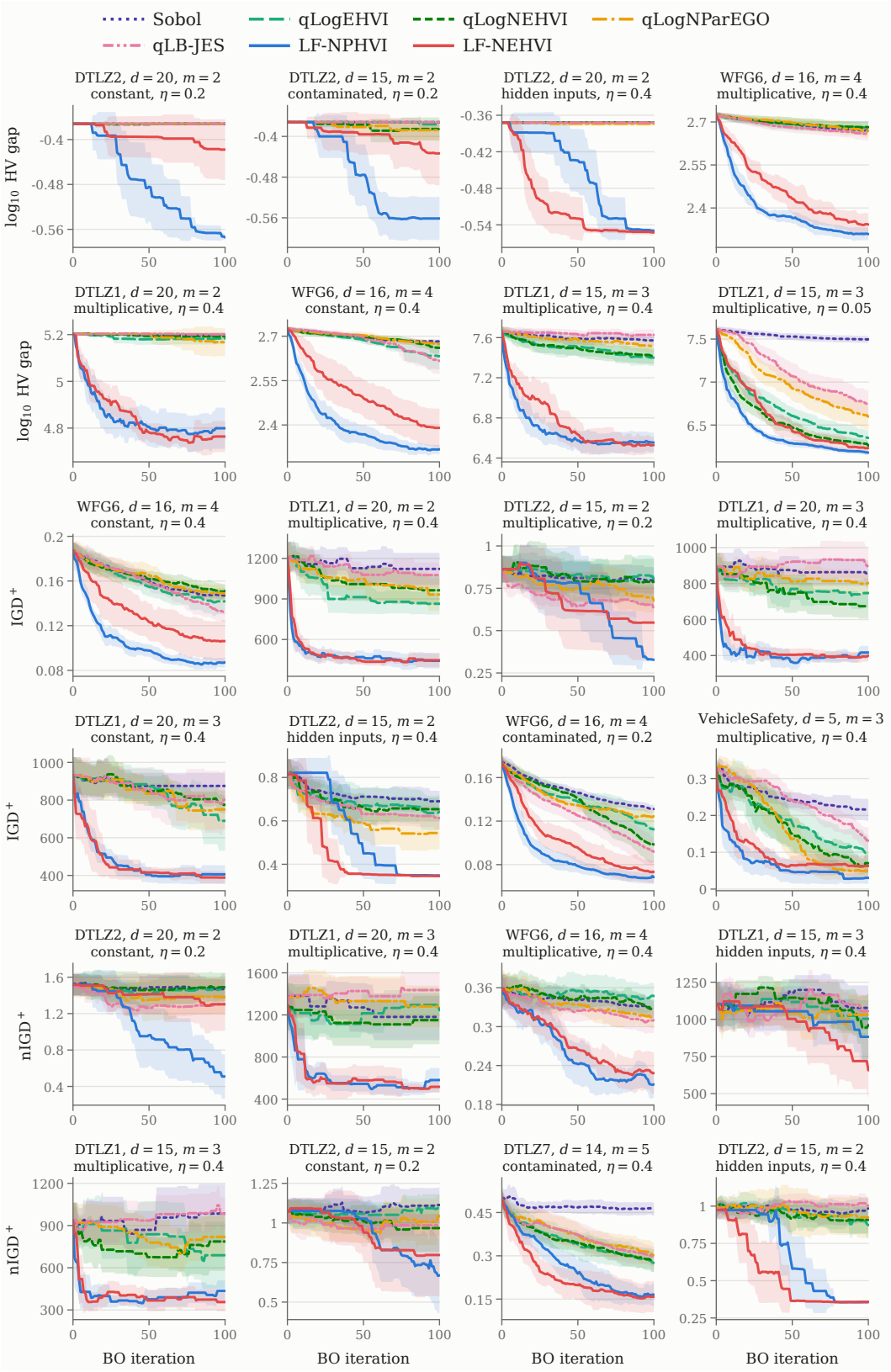


Figure 2: Convergence on the returned set of the likelihood-free estimator, on twenty-four conditions over eight problem instances, spanning $d = 5$ to 20, $m = 2$ to 5, all four noise structures and all three noise levels. Rows one and two are the HV gap, three and four IGD$^+$, five and six nIGD$^+$.

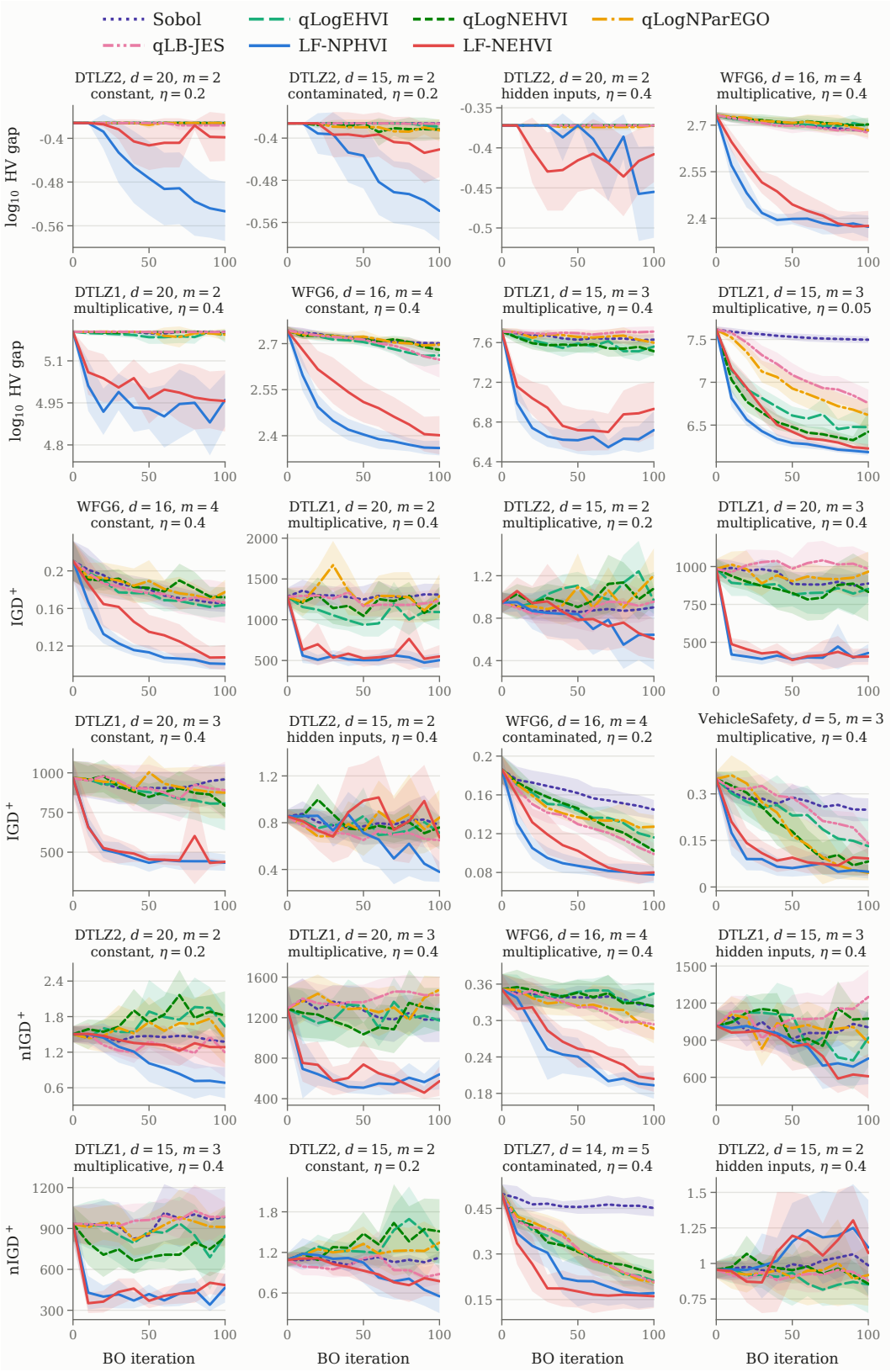


Figure 3: The same twenty-four conditions as Figure 2, in the same positions, scored on the returned set of the VHGP estimator, which is evaluated every ten iterations.

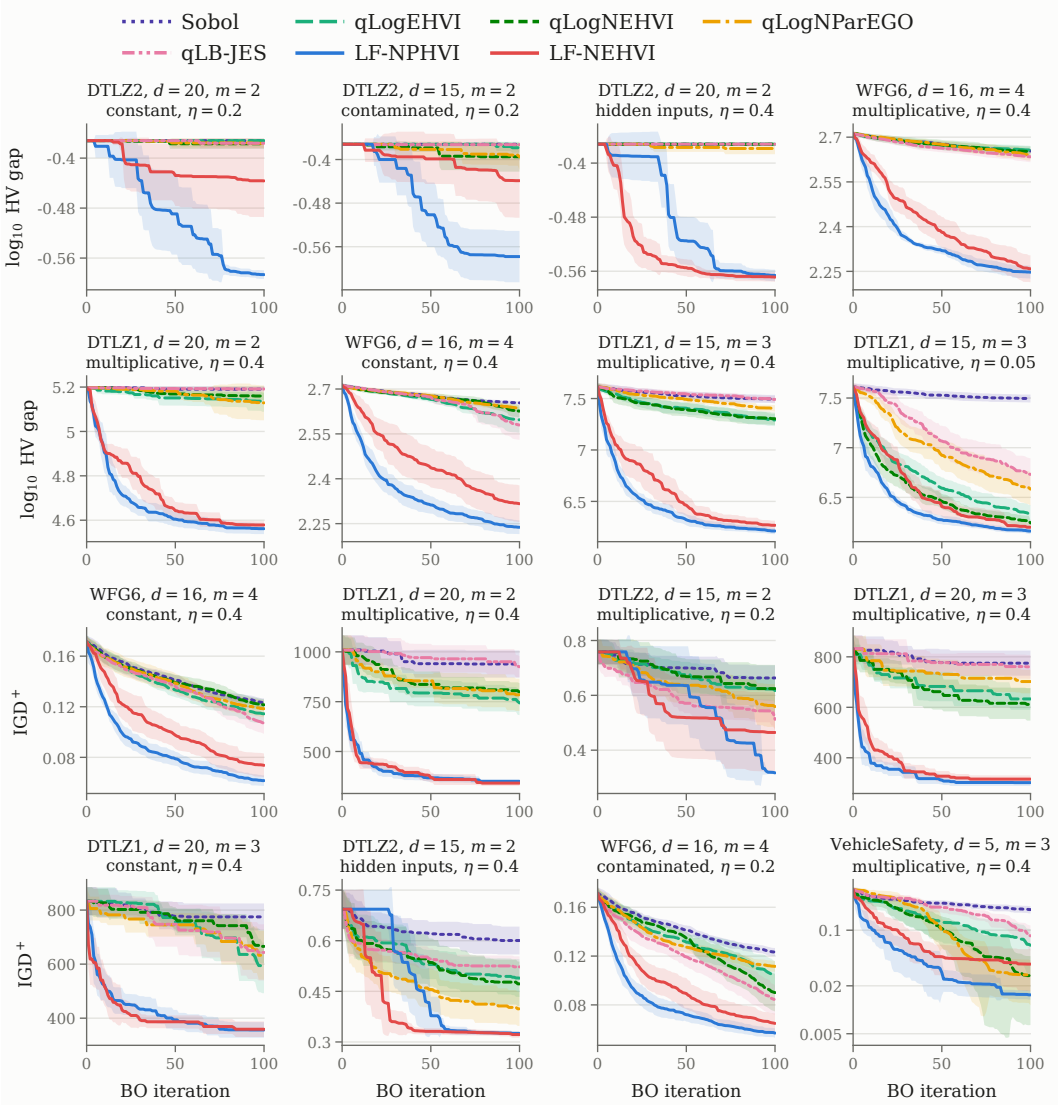


Figure 4: The first sixteen conditions of Figure 2, in the same positions, under the oracle values (11), which score every evaluated design. nIGD$^+$ needs an estimate of the objectives and is not defined for the oracle, so its two rows are absent.

### G.3 Ranks by noise structure

Tables 7 to 9 split the ranks by noise structure. On HV gap the crossover appears under all four, the homoscedastic Gaussian one included, so it is driven by the amount of noise and not by misspecification alone. The margin is smallest under contaminated noise, where at $\eta = 0.4$ a baseline ranks first on IGD$^+$ and nIGD$^+$.

Table 7: Mean rank on $\log_{10}$ HV gap by noise structure, likelihood-free estimator, over the 15 problems of each cell; lower is better, best per column in bold.

| | constant | | | contaminated | | | hidden inputs | | | multiplicative | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\eta$ | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 |
| Sobol | 6.70 | 6.77 | 6.73 | 6.63 | 6.73 | 6.33 | 6.80 | 6.60 | 6.60 | 6.80 | 6.47 | 5.93 |
| qLogEHVI | **2.00** | 3.83 | 4.13 | 2.33 | 4.50 | 4.50 | 2.60 | 4.13 | 3.57 | 2.60 | 4.20 | 4.27 |
| qLogNEHVI | 2.53 | 3.80 | 4.23 | **1.93** | 3.63 | 3.87 | **1.73** | 3.87 | 3.97 | **1.73** | 3.90 | 4.00 |
| qLogNParEGO | 4.67 | 4.63 | 4.53 | 4.20 | 4.27 | 4.83 | 3.80 | 4.20 | 3.87 | 4.47 | 4.47 | 3.93 |
| qLB-JES | 5.70 | 3.83 | 4.17 | 5.63 | 3.27 | 3.13 | 5.57 | 4.13 | 5.27 | 5.67 | 4.77 | 5.33 |
| LF-NPHVI | 3.07 | **2.33** | **2.07** | 3.20 | **2.27** | **2.53** | 2.67 | **2.33** | 2.60 | 2.93 | **2.07** | 2.40 |
| LF-NEHVI | 3.33 | 2.80 | 2.13 | 4.07 | 3.33 | 2.80 | 4.83 | 2.73 | **2.13** | 3.80 | 2.13 | **2.13** |

Table 8: Mean rank on IGD$^+$ by noise structure, likelihood-free estimator, over the 16 problems of each cell; lower is better, best per column in bold.

| | constant | | | contaminated | | | hidden inputs | | | multiplicative | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\eta$ | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 |
| Sobol | 6.38 | 6.50 | 6.19 | 6.31 | 6.50 | 6.00 | 6.38 | 6.38 | 6.19 | 6.50 | 6.06 | 5.62 |
| qLogEHVI | **2.25** | 4.31 | 3.88 | **2.31** | 4.25 | 4.31 | 3.19 | 4.56 | 3.56 | 2.88 | 4.38 | 3.75 |
| qLogNEHVI | 2.56 | 3.75 | 4.12 | 2.62 | 3.12 | 3.56 | **2.19** | 3.75 | 3.44 | **1.88** | 4.38 | 3.94 |
| qLogNParEGO | 2.75 | 4.12 | 4.56 | 2.50 | 3.56 | 4.31 | 2.50 | 3.25 | 3.69 | 3.19 | 3.44 | 3.75 |
| qLB-JES | 5.44 | 3.00 | 3.75 | 5.25 | 3.31 | **2.75** | 5.44 | 3.56 | 5.12 | 5.50 | 4.12 | 4.25 |
| LF-NPHVI | 3.75 | **2.81** | 2.81 | 3.75 | **2.81** | 3.19 | 3.06 | **2.94** | 3.38 | 3.50 | **2.75** | **3.31** |
| LF-NEHVI | 4.88 | 3.50 | **2.69** | 5.25 | 4.44 | 3.88 | 5.25 | 3.56 | **2.62** | 4.56 | 2.88 | 3.38 |

Table 9: Mean rank on nIGD$^+$ by noise structure, likelihood-free estimator, over the 16 problems of each cell; lower is better, best per column in bold.

| | constant | | | contaminated | | | hidden inputs | | | multiplicative | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\eta$ | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 | 0.05 | 0.2 | 0.4 |
| Sobol | 6.38 | 6.56 | 5.56 | 6.69 | 5.94 | 5.31 | 6.00 | 5.75 | 5.94 | 6.38 | 5.62 | 5.56 |
| qLogEHVI | **2.38** | 3.94 | 3.56 | 2.81 | 4.31 | 3.81 | 3.00 | 4.44 | 3.62 | 3.12 | 4.44 | 3.44 |
| qLogNEHVI | 2.69 | 4.19 | 3.62 | 2.75 | 3.88 | **3.31** | 2.56 | 3.62 | 3.50 | **2.12** | 4.25 | 4.12 |
| qLogNParEGO | 2.50 | 4.31 | 4.25 | **2.31** | 4.00 | 4.06 | **2.12** | **2.88** | 3.38 | 2.75 | 3.44 | 3.75 |
| qLB-JES | 5.25 | 3.56 | 4.75 | 5.12 | 3.94 | 4.25 | 5.25 | 4.00 | 5.88 | 5.56 | 4.50 | 4.25 |
| LF-NPHVI | 3.94 | 3.12 | 3.50 | 3.75 | **2.62** | 3.69 | 3.69 | 3.06 | **2.81** | 3.38 | 3.00 | 3.50 |
| LF-NEHVI | 4.88 | **2.31** | **2.75** | 4.56 | 3.31 | 3.56 | 5.38 | 4.25 | 2.88 | 4.69 | **2.75** | **3.38** |

### G.4 The loss decomposition

Table 10 reports the three terms of the decomposition (17) at the final iteration, each as a share of the decision maker's loss nIGD$^+$, together with the size of the returned set relative to the truly non-dominated evaluated designs.

Table 10: Decision-loss decomposition (17) at the final iteration, averaged over the seven methods, all problems and all noise structures, and the median size of the returned set relative to the truly non-dominated evaluated designs.

| estimator | $\eta$ | sampling | set selection | mis-selection | $\lvert\widehat{\mathcal{S}}_n\rvert/\lvert\mathcal{P}_n\rvert$ |
|---|---|---|---|---|---|
| LF | 0.05 | 0.72 | 0.02 | 0.26 | 1.35 |
| | 0.2 | 0.55 | 0.08 | 0.38 | 1.19 |
| | 0.4 | 0.47 | 0.12 | 0.42 | 1.05 |
| VHGP | 0.05 | 0.68 | 0.04 | 0.28 | 1.10 |
| | 0.2 | 0.57 | 0.13 | 0.31 | 0.80 |
| | 0.4 | 0.47 | 0.22 | 0.31 | 0.50 |

### G.5 The loss decomposition by problem group

Table 11 splits the decomposition of Table 10 by input-dimension group.

Table 11: Decision-loss decomposition (17) by input-dimension group (low: $d \leq 7$; high: $d \geq 8$), averaged over the seven methods. Identification is the larger share of the loss on the low-dimensional problems, where the sampled designs are already close to the front.

| estimator | group | $\eta$ | sampling | set selection | mis-selection | $\lvert\widehat{\mathcal{S}}_n\rvert/\lvert\mathcal{P}_n\rvert$ |
|---|---|---|---|---|---|---|
| likelihood-free | low | 0.05 | 0.52 | 0.01 | 0.47 | 1.14 |
| | low | 0.2 | 0.39 | 0.07 | 0.54 | 1.00 |
| | low | 0.4 | 0.33 | 0.10 | 0.57 | 1.00 |
| | high | 0.05 | 0.77 | 0.02 | 0.21 | 1.42 |
| | high | 0.2 | 0.58 | 0.08 | 0.34 | 1.26 |
| | high | 0.4 | 0.50 | 0.12 | 0.38 | 1.07 |
| VHGP | low | 0.05 | 0.47 | 0.06 | 0.47 | 0.97 |
| | low | 0.2 | 0.43 | 0.10 | 0.47 | 0.81 |
| | low | 0.4 | 0.35 | 0.20 | 0.45 | 0.55 |
| | high | 0.05 | 0.73 | 0.03 | 0.24 | 1.20 |
| | high | 0.2 | 0.60 | 0.13 | 0.27 | 0.79 |
| | high | 0.4 | 0.50 | 0.23 | 0.27 | 0.49 |

### G.6 Terminal values per problem

Tables 12 to 35 report the terminal value of every condition as the mean over ten seeds with its standard error: Tables 12 to 20 under the likelihood-free estimator, Tables 21 to 29 under the VHGP estimator, and Tables 30 to 35 under the oracle values (11), for which nIGD$^+$ is not defined. In every row the best mean is in bold, and † marks it together with every method not significantly different from it (two-sided Mann–Whitney $U$ test over seeds at the $5\%$ level).

Table 12: Terminal $\log_{10}$ HV gap at $\eta = 0.05$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, $5\%$); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.28 (0.009) | 1.62 (0.01) | **1.57 (0.01)**$^\dagger$ | 1.98 (0.03) | 2.00 (0.008) | 1.97 (0.03) | 1.95 (0.01) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.36 (0.03)$^\dagger$ | 4.37 (0.01)$^\dagger$ | 4.43 (0.03) | 4.43 (0.05)$^\dagger$ | **4.33 (0.02)**$^\dagger$ | **4.33 (0.03)**$^\dagger$ |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.37 (0.03)$^\dagger$ | 6.39 (0.06)$^\dagger$ | 6.64 (0.05) | 6.67 (0.06) | **6.34 (0.03)**$^\dagger$ | 6.36 (0.04)$^\dagger$ |
| DTLZ1 d20 m2 | 5.19 (0.006) | **4.61 (0.01)**$^\dagger$ | 4.63 (0.02)$^\dagger$ | 4.66 (0.02) | 4.71 (0.03) | 4.63 (0.02)$^\dagger$ | 4.63 (0.01)$^\dagger$ |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.85 (0.04)$^\dagger$ | 6.88 (0.03) | 7.09 (0.04) | 7.12 (0.03) | 6.78 (0.02)$^\dagger$ | **6.77 (0.03)**$^\dagger$ |
| DTLZ2 d15 m2 | -0.372 (0) | **-0.638 (0.02)**$^\dagger$ | -0.627 (0.02)$^\dagger$ | -0.529 (0.02) | -0.372 (0) | -0.580 (0.003) | -0.433 (0.03) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.176 (0.03)$^\dagger$ | **-0.191 (0.02)**$^\dagger$ | -0.124 (0.01) | -0.0932 (2e-4) | -0.188 (0.004)$^\dagger$ | -0.102 (0.009) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.505 (0.02) | -0.519 (0.02)$^\dagger$ | -0.450 (0.02) | -0.372 (0) | **-0.576 (0.002)**$^\dagger$ | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | **-0.151 (0.02)**$^\dagger$ | -0.121 (0.01)$^\dagger$ | -0.100 (0.003)$^\dagger$ | -0.0929 (0) | -0.128 (0.01)$^\dagger$ | -0.0929 (0) |

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| DTLZ2 m5 | 0.147 (0.003) | 0.0457 (0.02)† | **0.0374 (0.02)**† | 0.127 (0.009) | 0.153 (0.002) | 0.124 (0.003) | 0.112 (0.001) |
| Penicillin | 5.68 (0.004) | 5.60 (0.02)† | 5.61 (0.03)† | 5.60 (0.02)† | 5.62 (0.02)† | 5.59 (0.02)† | **5.58 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | 0.0532 (0.03) | **-0.113 (0.03)**† | 0.447 (0.03) | 0.297 (0.03) | 0.420 (0.06) | 0.393 (0.03) |
| WFG6 m4 | 2.65 (0.003) | **2.13 (0.009)**† | 2.15 (0.02)† | 2.40 (0.02) | 2.32 (0.03) | 2.21 (0.007) | 2.16 (0.01)† |
| ZDT1 d20 m2 | 1.41 (0.02) | **0.0969 (0.2)**† | 0.364 (0.09)† | 0.672 (0.05) | 0.620 (0.05) | 0.755 (0.06) | 0.426 (0.2)† |
| ZDT2 d20 m2 | 1.63 (0.01) | **-0.709 (0.2)**† | -0.565 (0.1)† | 0.0221 (0.3)† | 1.07 (0.05) | 1.07 (0.1) | 0.889 (0.2) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.61 (0.01)† | **1.58 (0.02)**† | 1.93 (0.03) | 2.02 (0.004) | 1.96 (0.01) | 1.95 (0.02) |
| DTLZ1 d15 m2 | 5.09 (0.02) | 4.44 (0.02) | 4.41 (0.04)† | 4.43 (0.03) | 4.46 (0.04) | **4.37 (0.06)**† | 4.42 (0.05)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.35 (0.02)† | 6.36 (0.01)† | 6.68 (0.07) | 6.63 (0.06) | **6.33 (0.03)**† | 6.38 (0.05)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | **4.62 (0.03)**† | 4.65 (0.02)† | 4.73 (0.02) | 4.69 (0.04)† | 4.65 (0.03)† | 4.64 (0.02)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.80 (0.04)† | **6.76 (0.03)**† | 7.05 (0.04) | 7.10 (0.05) | 6.80 (0.02)† | 6.87 (0.06)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.580 (0.03)† | **-0.606 (0.03)**† | -0.506 (0.03) | -0.372 (0) | -0.586 (0.002) | -0.372 (0) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | **-0.194 (0.02)**† | -0.186 (0.03)† | -0.147 (0.01)† | -0.0935 (3e-4) | -0.187 (0.002)† | -0.0930 (3e-5) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.495 (0.02) | -0.548 (0.02)† | -0.476 (0.02) | -0.372 (0) | **-0.579 (0.003)**† | -0.392 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.129 (0.02)† | -0.131 (0.02)† | -0.107 (0.007)† | -0.0929 (0) | **-0.147 (0.01)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003) | **0.0367 (0.03)**† | 0.0424 (0.02)† | 0.124 (0.01) | 0.154 (0.002) | 0.128 (0.007) | 0.108 (0.003)† |
| Penicillin | 5.68 (0.004) | 5.63 (0.01)† | **5.60 (0.02)**† | **5.60 (0.02)**† | 5.63 (0.02)† | **5.60 (0.02)**† | **5.60 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | 0.0705 (0.04) | **-0.101 (0.05)**† | 0.434 (0.03) | 0.335 (0.03) | 0.461 (0.06) | 0.440 (0.05) |
| WFG6 m4 | 2.65 (0.003) | 2.13 (0.01)† | **2.12 (0.01)**† | 2.43 (0.02) | 2.33 (0.02) | 2.20 (0.005) | 2.16 (0.008) |
| ZDT1 d20 m2 | 1.42 (0.02) | **0.0142 (0.2)**† | 0.177 (0.1)† | 0.587 (0.06) | 0.701 (0.03) | 0.712 (0.1) | 0.688 (0.1) |
| ZDT2 d20 m2 | 1.63 (0.01) | **-1.02 (0.05)**† | -0.729 (0.06) | 0.726 (0.2) | 1.14 (0.03) | 0.937 (0.2) | 0.916 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.28 (0.009) | 1.80 (0.008) | **1.75 (0.008)**† | 1.90 (0.01) | 2.07 (0.008) | 2.02 (0.01) | 1.97 (0.02) |
| DTLZ1 d15 m2 | 5.10 (0.02) | 4.58 (0.03) | 4.52 (0.02)† | 4.53 (0.02)† | 4.55 (0.02) | **4.47 (0.01)**† | 4.55 (0.02) |
| DTLZ1 d15 m3 | 7.53 (0.02) | **6.69 (0.03)**† | 6.70 (0.04)† | 6.83 (0.04) | 6.85 (0.04) | 6.72 (0.04)† | 6.79 (0.06)† |
| DTLZ1 d20 m2 | 5.20 (0.005) | 4.79 (0.01) | **4.73 (0.01)**† | 4.74 (0.02)† | 4.84 (0.04) | 4.77 (0.02)† | 4.80 (0.02) |
| DTLZ1 d20 m3 | 7.76 (0.008) | 7.15 (0.04) | 7.10 (0.04)† | 7.17 (0.04) | 7.43 (0.03) | **7.05 (0.02)**† | 7.22 (0.03) |
| DTLZ2 d15 m2 | -0.372 (0) | -0.558 (0.03) | **-0.628 (0.02)**† | -0.521 (0.02) | -0.372 (0) | -0.583 (0.003) | -0.392 (0.02) |
| DTLZ2 d15 m3 | -0.0966 (0.002) | **-0.194 (0.02)**† | **-0.194 (0.02)**† | -0.132 (0.01)† | -0.0973 (0.004) | -0.189 (0.003)† | -0.135 (0.01) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.408 (0.02) | -0.480 (0.03) | -0.434 (0.02) | -0.372 (0) | **-0.562 (0.002)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.135 (0.02)† | -0.107 (0.007) | -0.0970 (0.003) | -0.0932 (2e-4) | **-0.174 (0.002)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.140 (0.002) | **0.0869 (0.02)**† | 0.0990 (0.02)† | 0.134 (0.01) | 0.147 (0.003) | 0.118 (0.002)† | 0.0946 (0.006)† |
| Penicillin | 5.67 (0.003) | **5.47 (0.03)**† | 5.56 (0.02) | 5.55 (0.02) | 5.61 (0.01) | 5.56 (0.02) | 5.57 (0.02) |
| VehicleSafety | 1.24 (0.01) | 0.792 (0.003) | **0.765 (0.002)**† | 0.857 (0.007) | 0.814 (0.005) | 0.860 (0.009) | 0.869 (0.008) |
| WFG6 m4 | 2.65 (0.003) | 2.47 (0.006)† | **2.46 (0.007)**† | 2.55 (0.009) | 2.52 (0.009) | 2.48 (0.002) | 2.49 (0.006) |
| ZDT1 d20 m2 | 1.40 (0.02) | 0.659 (0.08) | **0.424 (0.05)**† | 0.711 (0.07) | 0.756 (0.03) | 0.616 (0.08) | 0.760 (0.06) |
| ZDT2 d20 m2 | 1.63 (0.01) | 1.12 (0.008) | **0.700 (0.03)**† | 0.966 (0.07) | 1.15 (0.04) | 0.912 (0.08) | 1.15 (0.08) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.78 (0.01) | **1.62 (0.02)**† | 1.94 (0.03) | 2.00 (0.009) | 1.97 (0.02) | 1.90 (0.008) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.31 (0.02) | 4.29 (0.02) | 4.30 (0.02)† | 4.41 (0.03) | **4.25 (0.003)**† | 4.27 (0.008)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.35 (0.04) | 6.26 (0.01) | 6.60 (0.08) | 6.74 (0.08) | **6.18 (0.02)**† | 6.23 (0.02)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.54 (0.01)† | **4.53 (0.01)**† | 4.54 (0.009)† | 4.67 (0.06) | 4.54 (0.01)† | 4.57 (0.01)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.75 (0.03)† | 6.70 (0.02)† | 6.97 (0.05) | 7.21 (0.05) | **6.68 (0.02)**† | 6.75 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.659 (0.03)† | **-0.716 (0.02)**† | -0.526 (0.03) | -0.372 (0) | -0.587 (0.003) | -0.372 (0) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.212 (0.03)† | **-0.271 (0.02)**† | -0.159 (0.01) | -0.0969 (0.003) | -0.191 (0.002) | -0.132 (0.02) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.560 (0.03)† | -0.548 (0.04)† | -0.434 (0.02) | -0.372 (0) | **-0.582 (0.003)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.129 (0.02) | -0.178 (0.02)† | -0.111 (0.009) | -0.0929 (0) | **-0.186 (0.002)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003)† | 0.102 (0.02)† | **0.0651 (0.03)**† | 0.146 (0.005) | 0.144 (0.005) | 0.131 (0.005)† | 0.0998 (0.008)† |
| Penicillin | 5.68 (0.004) | 5.60 (0.02)† | **5.57 (0.04)**† | 5.60 (0.02) | 5.64 (0.01) | 5.61 (0.02)† | **5.57 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | -0.0497 (0.02) | **-0.266 (0.03)**† | 0.426 (0.03) | 0.330 (0.03) | 0.462 (0.06) | 0.321 (0.03) |
| WFG6 m4 | 2.65 (0.003) | **2.13 (0.004)**† | **2.13 (0.01)**† | 2.42 (0.02) | 2.35 (0.009) | 2.20 (0.009) | 2.15 (0.01)† |
| ZDT1 d20 m2 | 1.41 (0.02) | 0.354 (0.05)† | 0.0733 (0.1)† | 0.673 (0.06) | 0.568 (0.03) | **-0.0113 (0.2)**† | 0.438 (0.1)† |
| ZDT2 d20 m2 | 1.63 (0.01) | -0.824 (0.09)† | **-0.947 (0.07)**† | 0.129 (0.3)† | 1.22 (0.02) | 0.609 (0.1) | 1.00 (0.1) |

Table 13: Terminal $\log_{10}$ HV gap at $\eta = 0.2$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.30 (0.008) | 2.11 (0.01) | **2.01 (0.03)**† | 2.19 (0.03) | 2.03 (0.02)† | **2.01 (0.02)**† | 2.08 (0.02) |
| DTLZ1 d15 m2 | 5.14 (0.02) | 4.53 (0.07)† | 4.72 (0.03) | 4.69 (0.04) | 4.70 (0.08)† | 4.56 (0.05)† | **4.52 (0.05)**† |
| DTLZ1 d15 m3 | 7.60 (0.02) | 7.09 (0.06) | 7.05 (0.05) | 7.27 (0.05) | 7.09 (0.05) | **6.63 (0.08)**† | 6.89 (0.09)† |
| DTLZ1 d20 m2 | 5.20 (5e-4) | **4.81 (0.05)**† | 4.98 (0.02) | 4.97 (0.04)† | 4.85 (0.03)† | 4.86 (0.05)† | 4.89 (0.03)† |
| DTLZ1 d20 m3 | 7.77 (0.007) | 7.48 (0.07) | 7.53 (0.06) | 7.62 (0.03) | 7.33 (0.05) | **7.07 (0.05)**† | 7.15 (0.06)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (0) | -0.375 (0.003) | -0.374 (0.002) | -0.372 (0) | **-0.571 (0.005)**† | -0.476 (0.03) |
| DTLZ2 d15 m3 | -0.0938 (9e-4) | -0.0957 (0.003) | -0.0972 (0.003) | -0.0952 (0.002) | -0.0950 (0.001) | **-0.143 (0.01)**† | -0.140 (0.02)† |

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (2e-5) | -0.372 (0) | -0.372 (0) | **-0.574 (0.002)**† | -0.418 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0933 (4e-4)† | -0.0933 (4e-4)† | **-0.126 (0.01)**† | -0.117 (0.01)† |
| DTLZ2 m5 | 0.149 (0.003)† | 0.147 (0.005) | 0.148 (0.005)† | 0.124 (0.008)† | 0.129 (0.008)† | 0.138 (0.006)† | **0.123 (0.008)**† |
| Penicillin | 5.70 (0.007)† | 5.67 (0.009)† | 5.66 (0.008)† | **5.65 (0.03)**† | 5.68 (0.01)† | 5.66 (0.02)† | 5.66 (0.02)† |
| VehicleSafety | 1.24 (0.02) | 0.615 (0.05) | 0.504 (0.04)† | 0.690 (0.03) | **0.429 (0.03)**† | 0.607 (0.06) | 0.716 (0.07) |
| WFG6 m4 | 2.67 (0.003) | 2.52 (0.03) | 2.55 (0.03) | 2.61 (0.004) | 2.45 (0.03) | **2.24 (0.006)**† | 2.33 (0.02) |
| ZDT1 d20 m2 | 1.49 (0.01) | **1.06 (0.05)**† | 1.08 (0.06)† | 1.30 (0.04) | 1.21 (0.08)† | 1.22 (0.04)† | 1.27 (0.05) |
| ZDT2 d20 m2 | 1.64 (0.005) | **1.16 (0.08)**† | 1.17 (0.06)† | 1.37 (0.04)† | 1.27 (0.04)† | 1.46 (0.03) | 1.40 (0.05)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.30 (0.01) | 2.05 (0.02)† | 2.02 (0.02)† | 2.12 (0.03) | 2.04 (0.02)† | **2.00 (0.01)**† | 2.08 (0.02) |
| DTLZ1 d15 m2 | 5.14 (0.01) | 4.73 (0.04) | 4.71 (0.05)† | 4.70 (0.06)† | 4.65 (0.06)† | 4.63 (0.05)† | **4.58 (0.05)**† |
| DTLZ1 d15 m3 | 7.59 (0.02) | 7.04 (0.08) | 6.82 (0.06)† | 7.14 (0.06) | **6.67 (0.07)**† | 6.72 (0.09)† | 6.78 (0.1)† |
| DTLZ1 d20 m2 | 5.20 (0.006) | 4.91 (0.04)† | 4.91 (0.03)† | 4.96 (0.04) | 4.85 (0.06)† | 4.85 (0.05)† | **4.81 (0.04)**† |
| DTLZ1 d20 m3 | 7.76 (0.01) | 7.48 (0.04) | 7.46 (0.08) | 7.59 (0.04) | 7.25 (0.09)† | **7.06 (0.04)**† | 7.19 (0.05)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.376 (0.004) | -0.386 (0.01) | -0.388 (0.01) | -0.372 (0) | **-0.561 (0.02)**† | -0.433 (0.03) |
| DTLZ2 d15 m3 | -0.0940 (9e-4) | -0.0965 (0.002) | -0.0934 (5e-4) | -0.0999 (0.007) | -0.0940 (9e-4) | **-0.151 (0.01)**† | -0.120 (0.01)† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.375 (0.003) | -0.372 (0) | **-0.547 (0.02)**† | -0.411 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0933 (4e-4)† | -0.0929 (0)† | -0.0937 (6e-4)† | -0.0947 (0.002)† | **-0.110 (0.01)**† | -0.0929 (0)† |
| DTLZ2 m5 | 0.148 (0.002) | 0.147 (0.005) | 0.146 (0.004) | 0.133 (0.006)† | **0.124 (0.007)**† | 0.140 (0.005)† | 0.125 (0.008)† |
| Penicillin | 5.70 (0.009) | 5.66 (0.01)† | **5.62 (0.02)**† | 5.63 (0.02)† | 5.67 (0.02)† | 5.63 (0.02)† | 5.66 (0.02)† |
| VehicleSafety | 1.23 (0.02) | 0.531 (0.05) | **0.360 (0.03)**† | 0.558 (0.04) | 0.484 (0.06)† | 0.610 (0.07) | 0.625 (0.07) |
| WFG6 m4 | 2.66 (0.003) | 2.50 (0.04) | 2.43 (0.04) | 2.59 (0.01) | 2.43 (0.04) | **2.24 (0.02)**† | 2.31 (0.02)† |
| ZDT1 d20 m2 | 1.49 (0.008) | 1.06 (0.07)† | **0.877 (0.09)**† | 1.19 (0.06) | 1.05 (0.1)† | 1.21 (0.06) | 1.35 (0.03) |
| ZDT2 d20 m2 | 1.65 (0.006) | 0.754 (0.2)† | **0.549 (0.2)**† | 1.26 (0.06) | 1.17 (0.06) | 1.45 (0.03) | 1.47 (0.02) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.29 (0.008) | 2.02 (0.01) | **1.94 (0.02)**† | 1.98 (0.02)† | 2.11 (0.01) | 2.03 (0.01) | 2.04 (0.02) |
| DTLZ1 d15 m2 | 5.15 (0.02) | 4.82 (0.04) | 4.75 (0.04) | 4.73 (0.04) | 4.90 (0.09) | **4.52 (0.01)**† | 4.55 (0.03)† |
| DTLZ1 d15 m3 | 7.66 (0.03) | 7.25 (0.05) | 7.25 (0.07) | 7.40 (0.05) | 7.27 (0.08) | **6.77 (0.05)**† | 6.82 (0.07)† |
| DTLZ1 d20 m2 | 5.20 (5e-4) | 4.95 (0.02) | 4.97 (0.05) | 4.95 (0.03) | 5.20 (0.004) | 4.83 (0.03)† | **4.82 (0.01)**† |
| DTLZ1 d20 m3 | 7.79 (0.005) | 7.60 (0.05) | 7.60 (0.04) | 7.65 (0.04) | 7.65 (0.06) | 7.22 (0.04)† | **7.11 (0.02)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.379 (0.005) | -0.372 (3e-4) | -0.377 (0.005) | -0.382 (0.01) | **-0.564 (0.004)**† | -0.503 (0.03)† |
| DTLZ2 d15 m3 | -0.0961 (0.002) | -0.0969 (0.002) | -0.0952 (0.002) | -0.0981 (0.003) | -0.0996 (0.003) | -0.166 (0.004) | **-0.182 (0.005)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.551 (0.001)**† | -0.411 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0934 (4e-4) | -0.0929 (0) | -0.0975 (0.003) | **-0.171 (0.002)**† | -0.167 (0.009)† |
| DTLZ2 m5 | 0.144 (0.001) | 0.147 (0.003) | 0.147 (0.006) | 0.125 (0.008)† | 0.129 (0.007)† | 0.141 (0.004) | **0.113 (0.004)**† |
| Penicillin | 5.67 (0.008) | **5.49 (0.02)**† | 5.55 (0.02)† | 5.57 (0.03)† | 5.57 (0.02) | 5.55 (0.02)† | 5.56 (0.02) |
| VehicleSafety | 1.25 (0.01) | 0.874 (0.01) | **0.805 (0.004)**† | 0.856 (0.01) | 0.829 (0.005) | 0.905 (0.02) | 0.869 (0.01) |
| WFG6 m4 | 2.70 (0.006) | 2.65 (0.01)† | 2.68 (0.01) | 2.66 (0.02)† | 2.64 (0.01)† | **2.63 (0.02)**† | 2.66 (0.007)† |
| ZDT1 d20 m2 | 1.44 (0.02) | 1.13 (0.04)† | **1.04 (0.06)**† | 1.28 (0.07) | 1.16 (0.04)† | 1.09 (0.08)† | 1.16 (0.06)† |
| ZDT2 d20 m2 | 1.64 (0.01) | 1.37 (0.05) | 1.27 (0.08)† | 1.34 (0.02) | **1.22 (0.03)**† | 1.34 (0.04) | 1.39 (0.03) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.31 (0.01) | 2.18 (0.02) | 2.13 (0.02) | 2.16 (0.03) | 2.13 (0.009) | **2.06 (0.02)**† | 2.11 (0.02)† |
| DTLZ1 d15 m2 | 5.15 (0.02) | 4.73 (0.05) | 4.69 (0.05) | 4.63 (0.07) | 5.07 (0.03) | **4.39 (0.01)**† | 4.42 (0.02)† |
| DTLZ1 d15 m3 | 7.55 (0.02) | 7.11 (0.06) | 7.10 (0.07) | 7.29 (0.04) | 7.51 (0.01) | **6.37 (0.02)**† | 6.43 (0.03)† |
| DTLZ1 d20 m2 | 5.20 (0.006) | 4.94 (0.05) | 4.96 (0.03) | 4.93 (0.05) | 5.19 (0.006) | **4.64 (0.02)**† | **4.64 (0.01)**† |
| DTLZ1 d20 m3 | 7.76 (0.01) | 7.47 (0.06) | 7.49 (0.05) | 7.65 (0.04) | 7.75 (0.01) | **6.85 (0.02)**† | 6.97 (0.04) |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (0) | -0.374 (0.002) | -0.372 (0) | -0.372 (0) | **-0.575 (0.002)**† | -0.505 (0.03)† |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.0950 (0.001) | -0.0932 (3e-4) | -0.0966 (0.004) | -0.0946 (0.001) | -0.123 (0.02)† | **-0.168 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.495 (0.03)**† | -0.392 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0931 (1e-4)† | -0.105 (0.009)† | **-0.129 (0.01)**† |
| DTLZ2 m5 | 0.150 (0.002) | 0.148 (0.004) | 0.151 (0.004) | 0.140 (0.009)† | 0.138 (0.007)† | 0.155 (0.002) | **0.124 (0.007)**† |
| Penicillin | 5.69 (0.005) | **5.60 (0.03)**† | 5.62 (0.02)† | 5.66 (0.01)† | 5.63 (0.02)† | 5.61 (0.02)† | 5.62 (0.02)† |
| VehicleSafety | 1.23 (0.02) | 0.618 (0.05)† | 0.457 (0.06)† | 0.682 (0.06) | **0.407 (0.05)**† | 0.586 (0.05) | 0.575 (0.06)† |
| WFG6 m4 | 2.66 (0.003) | 2.61 (0.01) | 2.61 (0.01) | 2.61 (0.008) | 2.64 (0.01) | **2.27 (0.02)**† | 2.28 (0.02)† |
| ZDT1 d20 m2 | 1.47 (0.03) | 1.04 (0.1) | 0.999 (0.1) | 1.23 (0.08) | 1.46 (0.02) | **0.630 (0.1)**† | 0.869 (0.2) |
| ZDT2 d20 m2 | 1.65 (0.02) | 0.999 (0.1)† | **0.791 (0.2)**† | 1.07 (0.07)† | 1.46 (0.04) | 1.20 (0.07) | 1.29 (0.05) |

Table 14: Terminal $\log_{10}$ HV gap at $\eta = 0.4$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.31 (0.009) | 2.24 (0.03) | 2.22 (0.02) | 2.26 (0.02) | 2.22 (0.03) | **2.09 (0.02)**† | 2.19 (0.01) |
| DTLZ1 d15 m2 | 5.15 (0.02) | 4.86 (0.05)† | 4.94 (0.05) | 4.99 (0.04) | 5.12 (0.04) | **4.69 (0.05)**† | 4.70 (0.06)† |
| DTLZ1 d15 m3 | 7.62 (0.01) | 7.35 (0.08) | 7.49 (0.03) | 7.48 (0.05) | 7.33 (0.08) | **6.81 (0.08)**† | 6.92 (0.09)† |
| DTLZ1 d20 m2 | 5.20 (5e-4) | 5.09 (0.04) | 5.15 (0.02) | 5.15 (0.03) | 5.13 (0.04) | **4.82 (0.04)**† | 4.88 (0.04)† |
| DTLZ1 d20 m3 | 7.78 (0.005) | 7.69 (0.04) | 7.76 (0.007) | 7.72 (0.03) | 7.69 (0.06) | 7.19 (0.06)† | **7.14 (0.05)**† |

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.501 (0.03)**$^\dagger$ | -0.489 (0.03)$^\dagger$ |
| DTLZ2 d15 m3 | -0.0938 (9e-4) | -0.0953 (0.001) | -0.0941 (0.001) | -0.0969 (0.002) | -0.0987 (0.003) | -0.121 (0.01)$^\dagger$ | **-0.144 (0.01)**$^\dagger$ |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (3e-4) | -0.372 (0) | **-0.472 (0.03)**$^\dagger$ | -0.417 (0.02)$^\dagger$ |
| DTLZ2 d20 m3 | -0.0929 (0)$^\dagger$ | -0.0929 (0)$^\dagger$ | -0.0931 (2e-4)$^\dagger$ | -0.0942 (0.001)$^\dagger$ | -0.0962 (0.002)$^\dagger$ | -0.0949 (0.002)$^\dagger$ | **-0.110 (0.009)**$^\dagger$ |
| DTLZ2 m5 | 0.151 (0.002) | 0.128 (0.008)$^\dagger$ | 0.139 (0.006) | **0.117 (0.006)**$^\dagger$ | 0.121 (0.009)$^\dagger$ | 0.143 (0.006) | 0.137 (0.008)$^\dagger$ |
| Penicillin | 5.72 (0.007) | 5.71 (0.01)$^\dagger$ | 5.70 (0.02)$^\dagger$ | **5.68 (0.01)**$^\dagger$ | 5.72 (0.01)$^\dagger$ | 5.70 (0.02)$^\dagger$ | 5.71 (0.03) |
| VehicleSafety | 1.30 (0.009) | 0.905 (0.08) | 0.700 (0.03)$^\dagger$ | 0.959 (0.07) | 0.831 (0.06) | **0.648 (0.04)**$^\dagger$ | 0.793 (0.07)$^\dagger$ |
| WFG6 m4 | 2.68 (0.003) | 2.63 (0.02) | 2.66 (0.01) | 2.67 (0.006) | 2.62 (0.02) | **2.32 (0.008)**$^\dagger$ | 2.39 (0.03) |
| ZDT1 d20 m2 | 1.52 (0.008) | 1.41 (0.04)$^\dagger$ | 1.39 (0.04)$^\dagger$ | 1.49 (0.02)$^\dagger$ | 1.52 (0.01) | 1.42 (0.04)$^\dagger$ | **1.34 (0.06)**$^\dagger$ |
| ZDT2 d20 m2 | 1.68 (0.007) | 1.51 (0.04)$^\dagger$ | **1.47 (0.04)**$^\dagger$ | 1.60 (0.02) | 1.55 (0.04)$^\dagger$ | 1.53 (0.04)$^\dagger$ | 1.48 (0.05)$^\dagger$ |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.31 (0.01) | 2.23 (0.03) | 2.25 (0.02) | 2.22 (0.03) | 2.17 (0.03) | **2.04 (0.02)**$^\dagger$ | 2.19 (0.02) |
| DTLZ1 d15 m2 | 5.15 (0.01) | 4.87 (0.06)$^\dagger$ | 4.95 (0.05) | 4.92 (0.06) | 4.80 (0.08)$^\dagger$ | **4.66 (0.06)**$^\dagger$ | **4.66 (0.05)**$^\dagger$ |
| DTLZ1 d15 m3 | 7.61 (0.02) | 7.27 (0.07) | 7.27 (0.06) | 7.36 (0.03) | **6.76 (0.09)**$^\dagger$ | 6.85 (0.08)$^\dagger$ | 6.89 (0.1)$^\dagger$ |
| DTLZ1 d20 m2 | 5.20 (0.006) | 5.01 (0.05)$^\dagger$ | 5.05 (0.04) | 5.09 (0.04) | 4.91 (0.06)$^\dagger$ | **4.90 (0.03)**$^\dagger$ | 4.91 (0.05)$^\dagger$ |
| DTLZ1 d20 m3 | 7.77 (0.007) | 7.60 (0.06) | 7.64 (0.04) | 7.73 (0.02) | 7.39 (0.08)$^\dagger$ | 7.24 (0.05)$^\dagger$ | **7.21 (0.06)**$^\dagger$ |
| DTLZ2 d15 m2 | -0.372 (0)$^\dagger$ | -0.372 (0)$^\dagger$ | -0.372 (0)$^\dagger$ | -0.372 (0)$^\dagger$ | -0.372 (0)$^\dagger$ | -0.424 (0.03)$^\dagger$ | **-0.430 (0.03)**$^\dagger$ |
| DTLZ2 d15 m3 | -0.0940 (9e-4)$^\dagger$ | -0.0929 (0) | -0.0958 (0.002)$^\dagger$ | -0.0929 (0) | -0.0936 (6e-4)$^\dagger$ | -0.118 (0.01)$^\dagger$ | **-0.130 (0.01)**$^\dagger$ |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.452 (0.03)**$^\dagger$ | -0.423 (0.03)$^\dagger$ |
| DTLZ2 d20 m3 | -0.0929 (0)$^\dagger$ | -0.0929 (0)$^\dagger$ | -0.0931 (2e-4)$^\dagger$ | -0.0929 (0)$^\dagger$ | -0.0935 (6e-4)$^\dagger$ | **-0.113 (0.01)**$^\dagger$ | -0.102 (0.009)$^\dagger$ |
| DTLZ2 m5 | 0.149 (0.003)$^\dagger$ | 0.142 (0.005)$^\dagger$ | 0.144 (0.006)$^\dagger$ | **0.130 (0.007)**$^\dagger$ | 0.136 (0.004)$^\dagger$ | 0.152 (0.004) | 0.141 (0.007)$^\dagger$ |
| Penicillin | 5.72 (0.01) | 5.69 (0.02)$^\dagger$ | 5.67 (0.02)$^\dagger$ | **5.65 (0.02)**$^\dagger$ | 5.69 (0.01)$^\dagger$ | 5.69 (0.03)$^\dagger$ | 5.72 (0.03)$^\dagger$ |
| VehicleSafety | 1.27 (0.01) | 0.787 (0.04)$^\dagger$ | 0.681 (0.07)$^\dagger$ | 0.863 (0.04) | 0.786 (0.08)$^\dagger$ | 0.737 (0.07)$^\dagger$ | **0.662 (0.06)**$^\dagger$ |
| WFG6 m4 | 2.68 (0.003) | 2.66 (0.01) | 2.62 (0.02) | 2.64 (0.01) | 2.46 (0.04) | **2.29 (0.02)**$^\dagger$ | 2.33 (0.02)$^\dagger$ |
| ZDT1 d20 m2 | 1.51 (0.009) | 1.36 (0.06) | **1.23 (0.01)**$^\dagger$ | 1.46 (0.02) | 1.39 (0.05) | 1.41 (0.05) | 1.45 (0.04) |
| ZDT2 d20 m2 | 1.67 (0.008) | 1.44 (0.04)$^\dagger$ | **1.29 (0.07)**$^\dagger$ | 1.48 (0.02) | 1.42 (0.04) | 1.53 (0.02) | 1.57 (0.02) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.29 (0.008) | 2.09 (0.01) | **2.03 (0.01)**$^\dagger$ | 2.07 (0.02)$^\dagger$ | 2.15 (0.02) | 2.06 (0.01)$^\dagger$ | 2.13 (0.02) |
| DTLZ1 d15 m2 | 5.15 (0.02) | 4.76 (0.06) | 4.89 (0.04) | 4.95 (0.06) | 5.13 (0.02) | 4.56 (0.02)$^\dagger$ | **4.54 (0.03)**$^\dagger$ |
| DTLZ1 d15 m3 | 7.64 (0.03) | 7.35 (0.07) | 7.40 (0.06) | 7.48 (0.04) | 7.57 (0.09) | 6.85 (0.06)$^\dagger$ | **6.77 (0.06)**$^\dagger$ |
| DTLZ1 d20 m2 | 5.20 (5e-4) | 5.03 (0.04) | 5.12 (0.03) | 5.04 (0.04) | 5.20 (0.001) | 4.82 (0.04)$^\dagger$ | **4.79 (0.03)**$^\dagger$ |
| DTLZ1 d20 m3 | 7.79 (0.006) | 7.61 (0.04) | 7.67 (0.03) | 7.72 (0.03) | 7.73 (0.04) | 7.27 (0.05)$^\dagger$ | **7.15 (0.04)**$^\dagger$ |
| DTLZ2 d15 m2 | -0.372 (0) | -0.375 (0.003) | -0.373 (1e-3) | -0.376 (0.003) | -0.372 (0) | -0.551 (0.002)$^\dagger$ | **-0.554 (0.002)**$^\dagger$ |
| DTLZ2 d15 m3 | -0.0933 (3e-4) | -0.108 (0.006) | -0.0955 (0.002) | -0.111 (0.007) | -0.0969 (0.002) | -0.158 (0.008) | **-0.176 (0.001)**$^\dagger$ |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.374 (0.002) | -0.372 (0) | -0.550 (0.002)$^\dagger$ | **-0.553 (0.003)**$^\dagger$ |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0972 (0.003) | -0.0932 (2e-4) | -0.0933 (4e-4) | -0.0932 (3e-4) | -0.108 (0.01) | **-0.176 (0.002)**$^\dagger$ |
| DTLZ2 m5 | 0.150 (0.002) | 0.132 (0.007)$^\dagger$ | 0.132 (0.006)$^\dagger$ | 0.148 (0.004) | 0.134 (0.004) | 0.154 (0.004) | **0.119 (0.004)**$^\dagger$ |
| Penicillin | 5.67 (0.007) | 5.57 (0.02)$^\dagger$ | 5.58 (0.02)$^\dagger$ | 5.57 (0.01)$^\dagger$ | 5.58 (0.02)$^\dagger$ | **5.55 (0.01)**$^\dagger$ | **5.55 (0.02)**$^\dagger$ |
| VehicleSafety | 1.26 (0.01) | 0.896 (0.006) | **0.861 (0.007)**$^\dagger$ | 0.886 (0.01)$^\dagger$ | 0.906 (0.03)$^\dagger$ | 0.923 (0.02) | 0.893 (0.005) |
| WFG6 m4 | 2.75 (0.007)$^\dagger$ | 2.76 (0.01)$^\dagger$ | 2.75 (0.008)$^\dagger$ | **2.73 (0.02)**$^\dagger$ | **2.73 (0.02)**$^\dagger$ | **2.73 (0.02)**$^\dagger$ | **2.73 (0.02)**$^\dagger$ |
| ZDT1 d20 m2 | 1.47 (0.03) | 1.39 (0.03) | **1.26 (0.04)**$^\dagger$ | 1.54 (0.03) | 1.49 (0.04) | 1.38 (0.05)$^\dagger$ | 1.41 (0.04) |
| ZDT2 d20 m2 | 1.65 (0.01) | **1.48 (0.06)**$^\dagger$ | 1.53 (0.05)$^\dagger$ | 1.52 (0.04)$^\dagger$ | 1.54 (0.05)$^\dagger$ | 1.51 (0.03)$^\dagger$ | 1.53 (0.03)$^\dagger$ |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.32 (0.01) | 2.30 (0.02) | 2.26 (0.01) | 2.26 (0.02) | 2.30 (0.01) | **2.12 (0.02)**$^\dagger$ | 2.24 (0.02) |
| DTLZ1 d15 m2 | 5.16 (0.02) | 5.02 (0.03) | 4.97 (0.07) | 5.10 (0.03) | 5.16 (0.01) | **4.46 (0.01)**$^\dagger$ | 4.49 (0.03)$^\dagger$ |
| DTLZ1 d15 m3 | 7.57 (0.03) | 7.40 (0.04) | 7.42 (0.04) | 7.52 (0.03) | 7.63 (0.03) | 6.55 (0.05)$^\dagger$ | **6.53 (0.04)**$^\dagger$ |
| DTLZ1 d20 m2 | 5.20 (0.006) | 5.18 (0.01) | 5.19 (0.007) | 5.17 (0.03) | 5.20 (0.002) | 4.80 (0.04)$^\dagger$ | **4.76 (0.03)**$^\dagger$ |
| DTLZ1 d20 m3 | 7.77 (0.01) | 7.72 (0.02) | 7.69 (0.02) | 7.76 (0.009) | 7.78 (0.01) | **7.09 (0.04)**$^\dagger$ | **7.09 (0.05)**$^\dagger$ |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.473 (0.03)**$^\dagger$ |
| DTLZ2 d15 m3 | -0.0929 (0) | -0.0932 (3e-4)$^\dagger$ | -0.0936 (6e-4)$^\dagger$ | -0.0942 (0.001)$^\dagger$ | -0.0931 (2e-4)$^\dagger$ | -0.0983 (0.005)$^\dagger$ | **-0.115 (0.01)**$^\dagger$ |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.499 (0.03)**$^\dagger$ |
| DTLZ2 d20 m3 | -0.0929 (0)$^\dagger$ | -0.0929 (0)$^\dagger$ | -0.0929 (0)$^\dagger$ | -0.0952 (0.002)$^\dagger$ | -0.0929 (0)$^\dagger$ | -0.0929 (0)$^\dagger$ | **-0.111 (0.01)**$^\dagger$ |
| DTLZ2 m5 | 0.152 (0.002)$^\dagger$ | 0.154 (0.002)$^\dagger$ | 0.150 (0.005)$^\dagger$ | **0.127 (0.01)**$^\dagger$ | 0.140 (0.006)$^\dagger$ | 0.159 (9e-4) | 0.159 (9e-4) |
| Penicillin | 5.70 (0.009) | 5.64 (0.01)$^\dagger$ | 5.65 (0.006)$^\dagger$ | 5.68 (0.02) | 5.68 (0.01) | **5.61 (0.02)**$^\dagger$ | 5.68 (0.02) |
| VehicleSafety | 1.26 (0.02) | 0.980 (0.07) | 0.890 (0.05) | 0.846 (0.04) | 1.08 (0.07) | **0.629 (0.07)**$^\dagger$ | 0.879 (0.07) |
| WFG6 m4 | 2.67 (0.003) | 2.68 (0.009) | 2.68 (0.01) | 2.67 (0.01) | 2.66 (0.01) | **2.31 (0.01)**$^\dagger$ | 2.34 (0.02)$^\dagger$ |
| ZDT1 d20 m2 | 1.48 (0.02) | 1.39 (0.07) | 1.42 (0.04) | 1.50 (0.04) | 1.48 (0.03) | **1.21 (0.04)**$^\dagger$ | 1.22 (0.09)$^\dagger$ |
| ZDT2 d20 m2 | 1.66 (0.02) | 1.57 (0.04)$^\dagger$ | 1.55 (0.05)$^\dagger$ | 1.62 (0.03) | 1.64 (0.01) | **1.30 (0.2)**$^\dagger$ | 1.52 (0.06)$^\dagger$ |

Table 15: Terminal $\mathrm{IGD}^+$ at $\eta = 0.05$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.105 (0.003) | 0.0194 (8e-4)$^\dagger$ | **0.0171 (0.001)**$^\dagger$ | 0.0503 (0.006) | 0.0639 (0.002) | 0.0385 (0.002) | 0.0614 (0.003) |
| DTLZ1 d15 m2 | 646 (30) | 263 (7) | 272 (7) | **240 (4)**$^\dagger$ | 269 (9) | 253 (6)$^\dagger$ | 271 (10)$^\dagger$ |
| DTLZ1 d15 m3 | 512 (30) | 225 (6)$^\dagger$ | 226 (5) | **209 (6)**$^\dagger$ | 267 (10) | 223 (7)$^\dagger$ | 222 (5)$^\dagger$ |
| DTLZ1 d20 m2 | 941 (30) | 352 (3) | 339 (6)$^\dagger$ | **338 (4)**$^\dagger$ | 407 (20) | 380 (10) | 369 (10) |

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| DTLZ1 d20 m3 | 775 (20) | 300 (8)† | **292 (7)**† | 309 (10)† | 361 (10) | 308 (10)† | 313 (7)† |
| DTLZ2 d15 m2 | 0.669 (0.02) | 0.207 (0.01)† | 0.238 (0.02)† | **0.194 (0.01)**† | 0.746 (0.02) | 0.324 (0.003) | 0.636 (0.07) |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.523 (0.07)† | 0.470 (0.06)† | 0.521 (0.05) | 0.752 (0.02) | **0.395 (0.007)**† | 0.761 (0.04) |
| DTLZ2 d20 m2 | 0.994 (0.02) | 0.354 (0.03)† | 0.405 (0.07)† | 0.338 (0.03)† | 1.04 (0.03) | **0.328 (0.002)**† | 1.08 (0.02) |
| DTLZ2 d20 m3 | 0.988 (0.02)† | **0.652 (0.1)**† | 0.781 (0.09)† | 0.683 (0.05)† | 1.01 (0.03) | 0.743 (0.1)† | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.482 (0.05)† | **0.463 (0.04)**† | 0.564 (0.03) | 0.696 (0.02) | 0.526 (0.01) | 0.475 (0.003) |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 0.00226 (7e-4) | 1.28e-4 (1e-4)† | 3.53e-5 (2e-5) |
| Penicillin | 0.0705 (0.001) | 0.0563 (0.006)† | 0.0664 (0.004)† | **0.0556 (0.002)**† | 0.0667 (0.002) | 0.0831 (0.004) | 0.0863 (0.002) |
| VehicleSafety | 0.182 (0.01) | 0.00547 (9e-4) | **0.00238 (3e-4)**† | 0.0150 (0.001) | 0.00991 (0.001) | 0.0149 (0.002) | 0.0158 (0.001) |
| WFG6 m4 | 0.124 (0.002) | **0.0370 (0.002)**† | 0.0414 (0.003)† | 0.0870 (0.003) | 0.0726 (0.005) | 0.0565 (0.002) | 0.0465 (0.003) |
| ZDT1 d20 m2 | 2.05 (0.09) | **0.115 (0.02)**† | 0.157 (0.03)† | 0.216 (0.03) | 0.375 (0.01) | 0.467 (0.06) | 0.422 (0.08) |
| ZDT2 d20 m2 | 3.23 (0.08) | **0.0884 (0.03)**† | 0.0971 (0.01)† | 0.154 (0.04)† | 0.751 (0.07) | 0.767 (0.1) | 0.652 (0.1) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.105 (0.003) | 0.0195 (8e-4)† | **0.0186 (0.002)**† | 0.0493 (0.006) | 0.0677 (0.001) | 0.0429 (0.004) | 0.0563 (0.003) |
| DTLZ1 d15 m2 | 647 (30) | 279 (8) | 270 (8)† | **248 (3)**† | 295 (10) | 268 (10)† | 292 (10) |
| DTLZ1 d15 m3 | 514 (30) | 216 (5)† | 234 (5) | **205 (6)**† | 259 (10) | 227 (10)† | 228 (7) |
| DTLZ1 d20 m2 | 949 (40) | **341 (5)**† | 364 (10)† | 349 (6)† | 356 (10)† | 382 (20)† | 382 (10) |
| DTLZ1 d20 m3 | 775 (20) | 293 (7)† | **285 (8)**† | 300 (10)† | 368 (10) | 326 (8) | 326 (10) |
| DTLZ2 d15 m2 | 0.669 (0.02) | 0.276 (0.03)† | 0.275 (0.03)† | **0.237 (0.03)**† | 0.736 (0.02) | 0.320 (0.002)† | 0.758 (0.02) |
| DTLZ2 d15 m3 | 0.695 (0.02) | 0.476 (0.06)† | 0.479 (0.06)† | 0.443 (0.04)† | 0.757 (0.02) | **0.397 (0.004)**† | 0.790 (0.02) |
| DTLZ2 d20 m2 | 0.990 (0.02) | 0.366 (0.03) | 0.336 (0.04)† | **0.291 (0.04)**† | 1.03 (0.03) | 0.325 (0.002) | 1.02 (0.08) |
| DTLZ2 d20 m3 | 0.983 (0.02)† | 0.766 (0.1)† | 0.782 (0.1)† | 0.685 (0.06)† | 1.05 (0.02) | **0.633 (0.1)**† | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.484 (0.05)† | 0.485 (0.05)† | 0.543 (0.03)† | 0.695 (0.02) | 0.568 (0.04) | **0.465 (0.006)**† |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 9.95e-4 (4e-4) | 3.02e-5 (2e-5) | 5.72e-5 (3e-5) |
| Penicillin | 0.0709 (0.001) | 0.0667 (0.002) | 0.0654 (0.003) | **0.0541 (0.002)**† | 0.0616 (0.004) | 0.0787 (0.004) | 0.0884 (0.004) |
| VehicleSafety | 0.182 (0.009) | 0.00555 (0.001) | **0.00298 (0.001)**† | 0.0150 (0.001) | 0.00997 (9e-4) | 0.0195 (0.005) | 0.0165 (0.001) |
| WFG6 m4 | 0.124 (0.002) | **0.0340 (0.002)**† | 0.0348 (0.002)† | 0.0912 (0.004) | 0.0712 (0.006) | 0.0551 (0.002) | 0.0441 (0.002) |
| ZDT1 d20 m2 | 2.07 (0.1) | **0.0889 (0.01)**† | 0.0941 (0.007)† | 0.168 (0.03)† | 0.384 (0.02) | 0.464 (0.1) | 0.532 (0.09) |
| ZDT2 d20 m2 | 3.23 (0.08) | **0.0461 (0.005)**† | 0.0895 (0.01) | 0.270 (0.03) | 0.805 (0.06) | 0.588 (0.09) | 0.858 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.107 (0.002) | 0.0358 (0.002) | **0.0287 (0.001)**† | 0.0580 (0.004) | 0.0771 (0.001) | 0.0519 (0.002) | 0.0598 (0.003) |
| DTLZ1 d15 m2 | 658 (40) | 352 (20) | 322 (10)† | **301 (2)**† | 326 (10)† | 313 (7)† | 350 (20) |
| DTLZ1 d15 m3 | 532 (30) | 265 (8)† | 264 (7)† | **253 (7)**† | 299 (9) | 303 (20) | 308 (10) |
| DTLZ1 d20 m2 | 997 (30) | 435 (10) | 405 (5)† | **397 (5)**† | 455 (30) | 431 (10) | 457 (20) |
| DTLZ1 d20 m3 | 834 (30) | 382 (20) | 358 (10)† | **332 (9)**† | 447 (20) | 350 (10)† | 426 (20) |
| DTLZ2 d15 m2 | 0.612 (0.02) | 0.259 (0.03)† | 0.230 (0.02)† | **0.220 (0.02)**† | 0.682 (0.04) | 0.318 (0.003) | 0.676 (0.05) |
| DTLZ2 d15 m3 | 0.628 (0.02) | 0.427 (0.05)† | 0.441 (0.05)† | 0.445 (0.03)† | 0.645 (0.03) | **0.397 (0.004)**† | 0.577 (0.06) |
| DTLZ2 d20 m2 | 0.935 (0.02) | 0.594 (0.05) | 0.451 (0.07)† | 0.367 (0.04)† | 1.01 (0.03) | **0.337 (0.003)**† | 1.03 (0.02) |
| DTLZ2 d20 m3 | 0.948 (0.02) | 0.715 (0.1) | 0.848 (0.08) | 0.817 (0.05) | 0.965 (0.04) | **0.424 (0.004)**† | 1.07 (0.02) |
| DTLZ2 m5 | 0.608 (0.009) | 0.527 (0.04)† | 0.564 (0.04)† | 0.567 (0.03) | 0.616 (0.02) | 0.513 (0.008) | **0.451 (0.006)**† |
| DTLZ7 m5 | 0.206 (0.008) | 8.91e-6 (2e-6) | **0 (0)**† | 4.38e-7 (3e-7)† | 0.0180 (0.005) | 1.56e-6 (2e-6)† | 5.90e-5 (5e-5) |
| Penicillin | 0.0743 (8e-4) | 0.0544 (0.004)† | 0.0678 (0.002) | **0.0525 (0.002)**† | 0.0685 (0.002) | 0.0854 (0.003) | 0.0814 (0.002) |
| VehicleSafety | 0.189 (0.01) | 0.0200 (4e-4) | **0.0174 (2e-4)**† | 0.0294 (8e-4) | 0.0240 (9e-4) | 0.0332 (0.002) | 0.0326 (7e-4) |
| WFG6 m4 | 0.125 (0.002) | 0.0956 (0.004)† | **0.0921 (0.004)**† | 0.114 (0.002) | 0.112 (0.003) | 0.105 (0.001) | 0.109 (0.003) |
| ZDT1 d20 m2 | 1.98 (0.08) | 0.423 (0.04) | **0.202 (0.02)**† | 0.303 (0.04)† | 0.450 (0.03) | 0.399 (0.05) | 0.442 (0.05) |
| ZDT2 d20 m2 | 3.25 (0.08) | 0.595 (0.02) | **0.280 (0.008)**† | 0.454 (0.05) | 0.861 (0.07) | 0.550 (0.1) | 0.991 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.105 (0.003) | 0.0304 (0.002) | **0.0200 (0.002)**† | 0.0588 (0.004) | 0.0621 (0.002) | 0.0324 (0.001) | 0.0530 (0.003) |
| DTLZ1 d15 m2 | 646 (30) | 259 (10)† | 252 (10)† | 244 (2)† | 270 (8) | **243 (2)**† | **243 (3)**† |
| DTLZ1 d15 m3 | 512 (30) | 217 (7)† | 210 (5)† | 217 (9)† | 260 (10) | **201 (2)**† | 214 (6)† |
| DTLZ1 d20 m2 | 939 (30) | 339 (5) | 337 (4)† | **329 (2)**† | 369 (20) | 340 (7)† | 342 (5) |
| DTLZ1 d20 m3 | 775 (20) | 297 (8)† | **290 (5)**† | 294 (9)† | 374 (20) | 300 (9)† | 310 (10) |
| DTLZ2 d15 m2 | 0.678 (0.03) | 0.193 (0.03)† | **0.164 (0.008)**† | 0.218 (0.02) | 0.723 (0.02) | 0.319 (0.002) | 0.771 (0.03) |
| DTLZ2 d15 m3 | 0.693 (0.02) | 0.458 (0.07)† | **0.319 (0.02)**† | 0.426 (0.05)† | 0.680 (0.03) | 0.391 (0.003)† | 0.641 (0.07) |
| DTLZ2 d20 m2 | 0.988 (0.02) | 0.390 (0.1)† | 0.517 (0.1)† | 0.376 (0.05)† | 1.02 (0.05) | **0.323 (0.003)**† | 1.09 (0.02) |
| DTLZ2 d20 m3 | 0.983 (0.02) | 0.754 (0.1) | 0.587 (0.1)† | 0.670 (0.07) | 0.983 (0.03) | **0.400 (0.003)**† | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.582 (0.05)† | 0.533 (0.05)† | 0.608 (0.03) | 0.640 (0.03) | 0.597 (0.03) | **0.455 (0.009)**† |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 0.00285 (6e-4) | 2.37e-5 (2e-5)† | 3.79e-6 (4e-6)† |
| Penicillin | 0.0715 (0.001) | 0.0636 (0.005)† | 0.0667 (0.002) | **0.0564 (0.002)**† | 0.0643 (0.004)† | 0.0911 (0.004) | 0.0906 (0.004) |
| VehicleSafety | 0.181 (0.009) | 0.00333 (2e-4) | **0.00159 (2e-4)**† | 0.0144 (0.001) | 0.00996 (0.001) | 0.0149 (0.001) | 0.0132 (0.001) |
| WFG6 m4 | 0.124 (0.002) | 0.0409 (0.001)† | **0.0387 (0.002)**† | 0.0899 (0.004) | 0.0788 (0.003) | 0.0561 (0.002) | 0.0426 (0.001)† |
| ZDT1 d20 m2 | 2.02 (0.09) | 0.142 (0.01) | **0.0778 (0.007)**† | 0.219 (0.03) | 0.331 (0.01) | 0.231 (0.05) | 0.336 (0.05) |
| ZDT2 d20 m2 | 3.23 (0.08) | 0.0828 (0.02)† | **0.0568 (0.008)**† | 0.177 (0.05)† | 0.969 (0.07) | 0.350 (0.04) | 0.742 (0.1) |

Table 16: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.2$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.120 (0.003) | 0.0953 (0.006) | **0.0735 (0.006)**† | 0.106 (0.008) | 0.0923 (0.006) | 0.0791 (0.009)† | 0.102 (0.01) |
| DTLZ1 d15 m2 | 765 (40) | 324 (30)† | 312 (10)† | **308 (20)**† | 399 (40) | 321 (20)† | 319 (10)† |
| DTLZ1 d15 m3 | 564 (10) | 288 (20)† | 274 (10)† | 359 (20) | 263 (20)† | **243 (10)**† | 257 (10)† |
| DTLZ1 d20 m2 | 1091 (40) | **406 (20)**† | 499 (30) | 530 (50) | 466 (20)† | 457 (40)† | 456 (10)† |
| DTLZ1 d20 m3 | 838 (30) | 475 (40) | 533 (40) | 592 (40) | 433 (30) | **352 (10)**† | 376 (10)† |
| DTLZ2 d15 m2 | 0.782 (0.03) | 0.768 (0.03) | 0.726 (0.05) | 0.648 (0.04) | 0.656 (0.04) | **0.332 (0.004)**† | 0.554 (0.09)† |
| DTLZ2 d15 m3 | 0.739 (0.02)† | 0.817 (0.04) | 0.748 (0.04)† | 0.656 (0.04)† | 0.572 (0.02)† | **0.566 (0.06)**† | 0.601 (0.07)† |
| DTLZ2 d20 m2 | 1.11 (0.02) | 1.06 (0.08) | 1.04 (0.08) | 0.885 (0.08) | 0.801 (0.03) | **0.329 (0.002)**† | 0.955 (0.1) |
| DTLZ2 d20 m3 | 1.06 (0.02) | 1.07 (0.03) | 1.08 (0.03) | 0.805 (0.03)† | **0.756 (0.04)**† | 0.855 (0.1)† | 0.877 (0.1)† |
| DTLZ2 m5 | 0.660 (0.01) | 0.668 (0.03) | 0.667 (0.03) | 0.558 (0.03)† | **0.536 (0.02)**† | 0.636 (0.04)† | 0.558 (0.04)† |
| DTLZ7 m5 | 0.199 (0.009) | 0.00862 (0.006)† | 0.00306 (0.002)† | 0.0297 (0.009) | 0.0192 (0.005) | **8.46e-4 (5e-4)**† | 0.00335 (0.001)† |
| Penicillin | 0.0843 (0.003) | 0.0791 (0.004)† | 0.0773 (0.004)† | **0.0751 (0.003)**† | 0.0769 (0.004)† | 0.100 (0.007) | 0.0925 (0.004) |
| VehicleSafety | 0.205 (0.01) | 0.0271 (0.005)† | **0.0163 (0.002)**† | 0.0272 (0.004) | 0.0173 (1e-3)† | 0.0365 (0.009) | 0.0350 (0.007) |
| WFG6 m4 | 0.135 (0.002) | 0.116 (0.006) | 0.119 (0.005) | 0.126 (0.003) | 0.101 (0.006) | **0.0702 (0.002)**† | 0.0764 (0.004)† |
| ZDT1 d20 m2 | 2.50 (0.09) | **0.872 (0.1)**† | 0.916 (0.1)† | 1.56 (0.2) | 1.36 (0.2)† | 1.28 (0.1) | 1.54 (0.2) |
| ZDT2 d20 m2 | 3.34 (0.05) | 0.972 (0.2)† | **0.922 (0.2)**† | 1.58 (0.2) | 1.22 (0.1)† | 2.08 (0.2) | 1.80 (0.3) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.114 (0.003) | 0.0854 (0.005)† | 0.0771 (0.006)† | 0.0892 (0.006)† | 0.0886 (0.007)† | **0.0724 (0.006)**† | 0.104 (0.006) |
| DTLZ1 d15 m2 | 734 (30) | 342 (10)† | 345 (20)† | **304 (20)**† | 342 (20)† | 322 (10)† | 322 (10)† |
| DTLZ1 d15 m3 | 561 (30) | 268 (20)† | **238 (10)**† | 327 (20) | 250 (8)† | 252 (10)† | 265 (10)† |
| DTLZ1 d20 m2 | 1053 (50) | 463 (40)† | **431 (30)**† | 492 (30)† | 520 (90)† | 472 (20)† | 440 (20)† |
| DTLZ1 d20 m3 | 815 (30) | 462 (30) | 478 (50)† | 568 (40) | 394 (20)† | **371 (10)**† | 397 (10)† |
| DTLZ2 d15 m2 | 0.845 (0.04) | 0.808 (0.05) | 0.670 (0.06) | 0.670 (0.06) | 0.724 (0.06) | **0.384 (0.06)**† | 0.711 (0.09)† |
| DTLZ2 d15 m3 | 0.723 (0.02)† | 0.764 (0.04)† | 0.760 (0.02)† | 0.695 (0.05)† | 0.627 (0.03)† | **0.561 (0.07)**† | 0.702 (0.07)† |
| DTLZ2 d20 m2 | 1.20 (0.03) | 1.10 (0.05) | 0.942 (0.08) | 0.745 (0.05) | 0.872 (0.06) | **0.423 (0.09)**† | 1.07 (0.1) |
| DTLZ2 d20 m3 | 1.03 (0.02) | 1.08 (0.04) | 1.10 (0.03) | 0.791 (0.05)† | **0.713 (0.04)**† | 0.969 (0.09) | 1.10 (0.02) |
| DTLZ2 m5 | 0.652 (0.01) | 0.672 (0.02) | 0.644 (0.02) | 0.572 (0.02)† | **0.536 (0.02)**† | 0.639 (0.03) | 0.573 (0.04)† |
| DTLZ7 m5 | 0.199 (0.009) | 0.00283 (1e-3)† | **1.80e-4 (7e-5)**† | 0.0120 (0.006) | 0.0129 (0.003) | 4.20e-4 (2e-4)† | 0.00373 (0.002)† |
| Penicillin | 0.0842 (0.004) | 0.0707 (0.005)† | 0.0744 (0.004)† | **0.0666 (0.003)**† | 0.0772 (0.003) | 0.105 (0.009) | 0.107 (0.006) |
| VehicleSafety | 0.191 (0.01) | 0.0200 (0.001) | **0.0101 (0.002)**† | 0.0195 (0.002) | 0.0207 (0.003) | 0.0257 (0.004) | 0.0258 (0.004) |
| WFG6 m4 | 0.131 (0.002) | 0.112 (0.007) | 0.0985 (0.008) | 0.124 (0.004) | 0.0919 (0.006) | **0.0685 (0.003)**† | 0.0734 (0.005)† |
| ZDT1 d20 m2 | 2.48 (0.05) | 0.932 (0.2)† | **0.653 (0.1)**† | 1.21 (0.2)† | 1.01 (0.2)† | 1.36 (0.2) | 1.78 (0.2) |
| ZDT2 d20 m2 | 3.47 (0.06) | 0.604 (0.1)† | **0.420 (0.1)**† | 1.22 (0.2) | 0.994 (0.1) | 2.04 (0.2) | 2.14 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.111 (0.002) | 0.0943 (0.007) | **0.0714 (0.006)**† | 0.0854 (0.009)† | 0.0829 (0.002)† | 0.0937 (0.005) | 0.102 (0.009) |
| DTLZ1 d15 m2 | 773 (50) | 405 (20) | 388 (20) | 365 (20) | 525 (70) | **313 (6)**† | 345 (20)† |
| DTLZ1 d15 m3 | 618 (40) | 329 (20) | 345 (20)† | 408 (30) | 401 (30) | **286 (8)**† | 294 (10)† |
| DTLZ1 d20 m2 | 1098 (40) | 477 (20)† | 533 (40)† | 515 (20)† | 1024 (50) | 474 (20)† | **469 (20)**† |
| DTLZ1 d20 m3 | 931 (40) | 581 (40) | 568 (40) | 617 (60) | 704 (80) | 381 (10)† | **373 (20)**† |
| DTLZ2 d15 m2 | 0.641 (0.03) | 0.589 (0.04) | 0.578 (0.05) | 0.505 (0.05) | 0.497 (0.03) | **0.334 (0.006)**† | 0.503 (0.08)† |
| DTLZ2 d15 m3 | 0.657 (0.02) | 0.683 (0.03) | 0.673 (0.03) | 0.554 (0.02) | 0.506 (0.02) | 0.442 (0.01) | **0.413 (0.005)**† |
| DTLZ2 d20 m2 | 0.969 (0.03) | 0.935 (0.03) | 0.911 (0.05) | 0.757 (0.04) | 0.669 (0.03) | **0.348 (0.001)**† | 0.908 (0.1) |
| DTLZ2 d20 m3 | 0.979 (0.02) | 0.999 (0.03) | 0.957 (0.04) | 0.742 (0.04) | 0.653 (0.03) | **0.430 (0.005)**† | 0.492 (0.07)† |
| DTLZ2 m5 | 0.629 (0.01) | 0.618 (0.02) | 0.618 (0.02) | 0.508 (0.02)† | 0.526 (0.02)† | 0.617 (0.02) | **0.481 (0.008)**† |
| DTLZ7 m5 | 0.207 (0.008) | 6.63e-4 (2e-4) | **1.60e-4 (6e-5)**† | 2.29e-4 (1e-4)† | 0.0253 (0.005) | 3.05e-4 (2e-4)† | 0.00282 (0.002) |
| Penicillin | 0.0776 (0.002) | 0.0629 (0.003)† | 0.0725 (0.003) | **0.0610 (0.002)**† | 0.0716 (0.001) | 0.0840 (0.003) | 0.0849 (0.002) |
| VehicleSafety | 0.197 (0.01) | 0.0382 (0.002) | **0.0234 (9e-4)**† | 0.0292 (0.002) | 0.0295 (0.002) | 0.0450 (0.006) | 0.0351 (0.002) |
| WFG6 m4 | 0.167 (0.005)† | 0.187 (0.008)† | 0.194 (0.01)† | **0.165 (0.01)**† | 0.178 (0.01)† | 0.190 (0.01)† | 0.206 (0.006) |
| ZDT1 d20 m2 | 2.22 (0.1) | 0.996 (0.09)† | **0.861 (0.1)**† | 1.63 (0.2) | 1.07 (0.09)† | 1.08 (0.2)† | 1.17 (0.2)† |
| ZDT2 d20 m2 | 3.38 (0.1) | 1.61 (0.2)† | 1.41 (0.3)† | 1.40 (0.1) | **1.08 (0.1)**† | 1.48 (0.2)† | 1.70 (0.1) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.117 (0.003) | 0.102 (0.007)† | 0.0962 (0.008)† | 0.0953 (0.01)† | 0.0994 (0.004)† | **0.0752 (0.009)**† | 0.115 (0.009) |
| DTLZ1 d15 m2 | 749 (40) | 377 (30) | 346 (30) | 318 (30)† | 631 (50) | **272 (8)**† | 290 (10)† |
| DTLZ1 d15 m3 | 557 (40) | 320 (20) | 321 (20) | 382 (30) | 506 (30) | **225 (7)**† | 245 (10)† |
| DTLZ1 d20 m2 | 1090 (60) | 493 (40) | 495 (30) | 494 (50) | 960 (60) | 375 (20)† | **355 (10)**† |
| DTLZ1 d20 m3 | 829 (40) | 450 (30) | 528 (30) | 623 (30) | 764 (30) | **326 (10)**† | 334 (10)† |
| DTLZ2 d15 m2 | 0.797 (0.04) | 0.805 (0.06) | 0.788 (0.07) | 0.677 (0.04) | 0.624 (0.06) | **0.328 (0.002)**† | 0.547 (0.1)† |
| DTLZ2 d15 m3 | 0.738 (0.03) | 0.706 (0.05) | 0.747 (0.05) | 0.661 (0.05)† | 0.527 (0.02) | 0.741 (0.08)† | **0.488 (0.06)**† |
| DTLZ2 d20 m2 | 1.08 (0.02) | 1.14 (0.06) | 1.11 (0.05) | 0.980 (0.05) | 0.886 (0.05)† | **0.575 (0.1)**† | 1.06 (0.1) |
| DTLZ2 d20 m3 | 1.03 (0.02) | 1.10 (0.04) | 1.13 (0.03) | 0.875 (0.03) | **0.681 (0.03)**† | 1.02 (0.09) | 0.833 (0.1)† |
| DTLZ2 m5 | 0.656 (0.01) | 0.678 (0.04) | 0.667 (0.02) | 0.598 (0.04)† | **0.577 (0.03)**† | 0.762 (0.02) | 0.590 (0.06)† |
| DTLZ7 m5 | 0.199 (0.009) | 1.48e-4 (6e-5) | 5.64e-4 (3e-4) | 8.43e-5 (8e-5)† | 0.0215 (0.005) | **9.02e-7 (9e-7)**† | 0.00237 (0.002) |
| Penicillin | 0.0767 (0.002)† | **0.0641 (0.006)**† | 0.0702 (0.003)† | 0.0657 (0.003)† | 0.0696 (0.003)† | 0.0926 (0.003) | 0.0896 (0.003) |
| VehicleSafety | 0.195 (0.01) | 0.0267 (0.004)† | **0.0172 (0.003)**† | 0.0333 (0.009)† | 0.0175 (0.002)† | 0.0232 (0.002)† | 0.0216 (0.003)† |
| WFG6 m4 | 0.132 (0.002) | 0.135 (0.004) | 0.140 (0.005) | 0.125 (0.003) | 0.139 (0.008) | 0.0786 (0.006)† | **0.0768 (0.004)**† |
| ZDT1 d20 m2 | 2.40 (0.1) | 0.998 (0.2) | 0.961 (0.2) | 1.45 (0.2) | 2.34 (0.1) | **0.466 (0.06)**† | 0.820 (0.1) |
| ZDT2 d20 m2 | 3.49 (0.2) | 0.688 (0.1)† | **0.573 (0.1)**† | 0.674 (0.1)† | 2.09 (0.2) | 1.08 (0.2) | 1.29 (0.2) |

Table 17: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.4$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.125 (0.004) | 0.139 (0.009) | 0.130 (0.01) | 0.136 (0.007) | 0.123 (0.005) | **0.101 (0.007)**† | 0.133 (0.007) |
| DTLZ1 d15 m2 | 789 (50) | 388 (40)† | 466 (40) | 529 (40) | 725 (60) | **347 (20)**† | 363 (30)† |
| DTLZ1 d15 m3 | 587 (10) | 409 (40) | 473 (30) | 476 (40) | 384 (50)† | **283 (10)**† | 293 (10)† |
| DTLZ1 d20 m2 | 1151 (60) | 700 (80) | 836 (80) | 857 (90) | 885 (90) | **422 (30)**† | 452 (20)† |
| DTLZ1 d20 m3 | 875 (30) | 690 (80) | 775 (20) | 750 (50) | 802 (80) | 406 (20)† | **389 (20)**† |
| DTLZ2 d15 m2 | 0.865 (0.03) | 0.788 (0.05) | 0.773 (0.03) | 0.754 (0.03)† | 0.700 (0.06)† | **0.520 (0.09)**† | 0.530 (0.09)† |
| DTLZ2 d15 m3 | 0.784 (0.02)† | 0.710 (0.04)† | 0.754 (0.03)† | 0.695 (0.04)† | 0.641 (0.04)† | 0.719 (0.08)† | **0.580 (0.07)**† |
| DTLZ2 d20 m2 | 1.15 (0.02)† | 1.11 (0.05)† | 1.06 (0.07)† | 1.10 (0.09)† | 0.978 (0.04)† | **0.747 (0.2)**† | 0.806 (0.1)† |
| DTLZ2 d20 m3 | 1.11 (0.02) | 0.916 (0.05)† | 0.945 (0.05) | 0.841 (0.06)† | **0.768 (0.06)**† | 1.10 (0.05) | 0.956 (0.1)† |
| DTLZ2 m5 | 0.683 (0.01) | 0.603 (0.03)† | 0.628 (0.03) | **0.551 (0.02)**† | 0.558 (0.03)† | 0.704 (0.05) | 0.666 (0.05)† |
| DTLZ7 m5 | 0.204 (0.008) | 0.0543 (0.01) | 0.0744 (0.02) | 0.102 (0.02) | 0.0448 (0.02) | **0.00367 (0.001)**† | 0.0355 (0.02)† |
| Penicillin | 0.0913 (0.004)† | **0.0906 (0.006)**† | 0.0951 (0.005)† | 0.0931 (0.005)† | 0.0938 (0.007)† | 0.125 (0.02) | 0.116 (0.008) |
| VehicleSafety | 0.248 (0.01) | 0.0682 (0.01) | **0.0248 (0.002)**† | 0.101 (0.03) | 0.0633 (0.02) | 0.0271 (0.003)† | 0.0549 (0.01) |
| WFG6 m4 | 0.147 (0.002) | 0.142 (0.006) | 0.151 (0.003) | 0.149 (0.004) | 0.131 (0.004) | **0.0872 (0.002)**† | 0.106 (0.009) |
| ZDT1 d20 m2 | 2.65 (0.06) | 2.09 (0.2)† | 1.95 (0.2)† | 2.52 (0.1)† | 2.67 (0.1) | 2.17 (0.2)† | **1.83 (0.2)**† |
| ZDT2 d20 m2 | 3.72 (0.07) | 2.44 (0.2)† | **2.15 (0.2)**† | 3.05 (0.2) | 2.73 (0.3)† | 2.60 (0.3)† | 2.34 (0.3)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.121 (0.004) | 0.121 (0.007) | 0.126 (0.007) | 0.122 (0.01) | 0.108 (0.006)† | **0.0910 (0.007)**† | 0.116 (0.01)† |
| DTLZ1 d15 m2 | 762 (30) | 433 (50)† | 503 (60) | 460 (60)† | 429 (60)† | 339 (20)† | **318 (8)**† |
| DTLZ1 d15 m3 | 559 (30) | 358 (30) | 344 (30) | 409 (30) | 274 (20)† | 298 (10)† | **264 (10)**† |
| DTLZ1 d20 m2 | 1106 (50) | 603 (60) | 594 (50)† | 648 (50) | 558 (80)† | **476 (20)**† | 479 (30)† |
| DTLZ1 d20 m3 | 861 (30) | 579 (70)† | 596 (40) | 755 (40) | 499 (60)† | **405 (10)**† | 411 (20)† |
| DTLZ2 d15 m2 | 0.891 (0.04)† | 0.873 (0.06)† | 0.949 (0.07)† | **0.835 (0.09)**† | 0.949 (0.04)† | 0.980 (0.2)† | 1.02 (0.2)† |
| DTLZ2 d15 m3 | 0.776 (0.02)† | 0.831 (0.03)† | 0.733 (0.05)† | 0.744 (0.03)† | 0.695 (0.04)† | 0.785 (0.09)† | **0.665 (0.07)**† |
| DTLZ2 d20 m2 | 1.19 (0.03)† | 1.16 (0.06)† | 1.14 (0.06)† | 1.11 (0.07)† | 1.02 (0.1)† | **0.909 (0.2)**† | 1.28 (0.3)† |
| DTLZ2 d20 m3 | 1.09 (0.02) | 1.03 (0.03) | 0.902 (0.06)† | 0.951 (0.07)† | **0.770 (0.05)**† | 0.976 (0.1)† | 1.12 (0.08) |
| DTLZ2 m5 | 0.662 (0.01)† | 0.655 (0.03)† | 0.651 (0.03)† | **0.597 (0.02)**† | 0.608 (0.02)† | 0.742 (0.03) | 0.700 (0.05)† |
| DTLZ7 m5 | 0.200 (0.009) | 0.0480 (0.01) | 0.0780 (0.02) | 0.0810 (0.02) | 0.0398 (0.01) | **6.25e-4 (4e-4)**† | 0.0212 (0.01) |
| Penicillin | 0.0944 (0.005) | 0.0914 (0.003) | 0.0861 (0.004)† | **0.0752 (0.004)**† | 0.0922 (0.003) | 0.121 (0.02) | 0.133 (0.02) |
| VehicleSafety | 0.226 (0.009) | 0.0434 (0.007)† | **0.0351 (0.01)**† | 0.0582 (0.01)† | 0.0505 (0.02)† | 0.0382 (0.009)† | 0.0424 (0.02)† |
| WFG6 m4 | 0.144 (0.002) | 0.149 (0.005) | 0.135 (0.006) | 0.140 (0.003) | 0.107 (0.008) | **0.0786 (0.004)**† | 0.0792 (0.003)† |
| ZDT1 d20 m2 | 2.59 (0.07) | 1.95 (0.3) | **1.16 (0.07)**† | 2.29 (0.1) | 2.03 (0.3) | 2.16 (0.3) | 2.27 (0.2) |
| ZDT2 d20 m2 | 3.58 (0.08) | 2.00 (0.2)† | **1.40 (0.2)**† | 2.15 (0.1) | 1.89 (0.2) | 2.50 (0.1) | 2.82 (0.1) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.112 (0.002) | 0.131 (0.007) | **0.0963 (0.007)**† | 0.114 (0.007)† | 0.104 (0.005)† | 0.104 (0.005)† | 0.138 (0.01) |
| DTLZ1 d15 m2 | 784 (50) | 404 (40)† | 448 (30) | 532 (50) | 716 (40) | 343 (10)† | **337 (10)**† |
| DTLZ1 d15 m3 | 611 (40) | 367 (40)† | 399 (40) | 460 (20) | 586 (40) | 317 (10)† | **287 (10)**† |
| DTLZ1 d20 m2 | 1090 (40) | 598 (40) | 709 (50) | 601 (50) | 1094 (60) | 455 (30)† | **448 (20)**† |
| DTLZ1 d20 m3 | 922 (40) | 575 (50) | 629 (50) | 725 (40) | 832 (70) | 433 (30)† | **410 (20)**† |
| DTLZ2 d15 m2 | 0.690 (0.02) | 0.643 (0.04) | 0.654 (0.03) | 0.544 (0.04) | 0.613 (0.03) | 0.348 (0.002)† | **0.346 (0.002)**† |
| DTLZ2 d15 m3 | 0.748 (0.006) | 0.641 (0.04) | 0.684 (0.03) | 0.598 (0.03) | 0.661 (0.03) | 0.483 (0.04) | **0.417 (0.002)**† |
| DTLZ2 d20 m2 | 0.994 (0.03) | 0.775 (0.04) | 0.801 (0.05) | 0.698 (0.03) | 0.827 (0.04) | 0.349 (0.002)† | **0.346 (0.002)**† |
| DTLZ2 d20 m3 | 1.02 (0.02) | 0.879 (0.06) | 0.835 (0.04) | 0.810 (0.05) | 0.812 (0.02) | 0.937 (0.09) | **0.420 (0.003)**† |
| DTLZ2 m5 | 0.680 (0.02) | 0.613 (0.03) | 0.612 (0.02) | 0.647 (0.02) | 0.590 (0.02) | 0.759 (0.03) | **0.513 (0.03)**† |
| DTLZ7 m5 | 0.207 (0.008) | 0.0149 (0.007) | 0.00332 (0.001)† | 0.00479 (0.003)† | 0.0491 (0.009) | **6.81e-4 (2e-4)**† | 0.0296 (0.01) |
| Penicillin | 0.0824 (0.003)† | 0.0759 (0.004)† | **0.0736 (0.003)**† | 0.0762 (0.003)† | 0.0777 (0.003)† | 0.0870 (0.003) | 0.0804 (0.003)† |
| VehicleSafety | 0.200 (0.01) | 0.0401 (0.002) | **0.0307 (9e-4)**† | 0.0356 (0.003)† | 0.0465 (0.008) | 0.0463 (0.006) | 0.0381 (0.001) |
| WFG6 m4 | 0.224 (0.01)† | 0.269 (0.02)† | 0.264 (0.02)† | **0.221 (0.02)**† | 0.292 (0.02) | 0.293 (0.03)† | 0.282 (0.01) |
| ZDT1 d20 m2 | 2.36 (0.2) | 1.92 (0.1) | **1.44 (0.2)**† | 2.92 (0.2) | 2.60 (0.3) | 2.00 (0.2)† | 2.07 (0.2) |
| ZDT2 d20 m2 | 3.45 (0.1) | **2.34 (0.3)**† | 2.60 (0.3)† | 2.52 (0.3)† | 2.75 (0.3)† | 2.40 (0.2)† | 2.55 (0.2)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.124 (0.003)† | 0.143 (0.01) | 0.147 (0.008) | 0.128 (0.01)† | 0.132 (0.004)† | **0.108 (0.01)**† | 0.170 (0.01) |
| DTLZ1 d15 m2 | 761 (40) | 539 (50) | 537 (60) | 654 (50) | 731 (30) | **307 (10)**† | 315 (10)† |
| DTLZ1 d15 m3 | 578 (40) | 442 (40) | 458 (30) | 512 (40) | 623 (30) | 265 (10)† | **252 (10)**† |
| DTLZ1 d20 m2 | 1122 (60) | 864 (40) | 964 (50) | 932 (60) | 1061 (50) | **442 (30)**† | 447 (30)† |
| DTLZ1 d20 m3 | 853 (40) | 747 (40) | 673 (30) | 803 (30) | 897 (50) | 417 (20)† | **399 (20)**† |
| DTLZ2 d15 m2 | 0.840 (0.03) | 0.790 (0.06)† | 0.776 (0.05)† | **0.627 (0.07)**† | 0.703 (0.03)† | 1.11 (0.06) | 0.720 (0.1)† |
| DTLZ2 d15 m3 | 0.789 (0.01) | 0.724 (0.04)† | 0.697 (0.04)† | 0.657 (0.05)† | **0.632 (0.01)**† | 0.938 (0.08) | 0.859 (0.1)† |
| DTLZ2 d20 m2 | 1.13 (0.03)† | 0.976 (0.05)† | 1.02 (0.06)† | 0.849 (0.05)† | 1.01 (0.03)† | 1.39 (0.04) | **0.663 (0.2)**† |
| DTLZ2 d20 m3 | 1.05 (0.02) | 0.936 (0.03) | 0.986 (0.04) | **0.761 (0.05)**† | 0.791 (0.03)† | 1.38 (0.07) | 1.31 (0.2) |
| DTLZ2 m5 | 0.680 (0.01) | 0.665 (0.02)† | 0.671 (0.03)† | **0.587 (0.03)**† | 0.643 (0.04)† | 0.970 (0.07) | 0.918 (0.05) |
| DTLZ7 m5 | 0.202 (0.009) | 0.0331 (0.007) | 0.0591 (0.01) | 0.0948 (0.02) | 0.0672 (0.01) | **6.63e-4 (6e-4)**† | 0.0352 (0.02) |
| Penicillin | 0.0835 (0.003)† | 0.0780 (0.003)† | **0.0742 (0.004)**† | 0.0849 (0.004)† | **0.0742 (0.003)**† | 0.0887 (0.003) | 0.103 (0.008) |
| VehicleSafety | 0.215 (0.01) | 0.0943 (0.02) | 0.0702 (0.01) | 0.0495 (0.01) | 0.132 (0.02) | **0.0305 (0.008)**† | 0.0628 (0.01)† |
| WFG6 m4 | 0.138 (0.002) | 0.146 (0.006) | 0.155 (0.007) | 0.149 (0.008) | 0.131 (0.006) | **0.0867 (0.003)**† | 0.0975 (0.006)† |
| ZDT1 d20 m2 | 2.45 (0.1) | 2.14 (0.3) | 2.13 (0.2) | 2.63 (0.2) | 2.44 (0.2) | **1.25 (0.1)**† | 1.46 (0.2)† |
| ZDT2 d20 m2 | 3.57 (0.1) | 2.92 (0.3)† | 2.77 (0.3)† | 3.23 (0.2) | 3.41 (0.1) | **2.00 (0.4)**† | 2.68 (0.3)† |

Table 18: Terminal nIGD$^+$ at $\eta = 0.05$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.150 (0.003) | **0.0732 (0.002)**† | 0.0749 (0.003)† | 0.0990 (0.006) | 0.120 (0.002) | 0.0981 (0.004) | 0.121 (0.003) |
| DTLZ1 d15 m2 | 700 (30) | 304 (20)† | 334 (20)† | **286 (10)**† | 312 (20)† | 335 (30)† | 318 (10)† |
| DTLZ1 d15 m3 | 569 (50) | 294 (20)† | 285 (10)† | **250 (10)**† | 361 (30) | 281 (20)† | 291 (20)† |
| DTLZ1 d20 m2 | 985 (50) | **386 (20)**† | 391 (20)† | 409 (30)† | 481 (30) | 458 (30)† | 453 (10) |
| DTLZ1 d20 m3 | 856 (40) | 372 (20)† | 389 (20)† | **359 (10)**† | 465 (20) | 382 (30)† | 422 (20) |
| DTLZ2 d15 m2 | 0.740 (0.03) | **0.249 (0.01)**† | 0.316 (0.02) | 0.277 (0.02)† | 0.818 (0.05) | 0.356 (0.008) | 0.693 (0.09) |
| DTLZ2 d15 m3 | 0.762 (0.02) | 0.596 (0.07)† | 0.551 (0.06) | 0.602 (0.05) | 0.807 (0.03) | **0.436 (0.01)**† | 0.812 (0.05) |
| DTLZ2 d20 m2 | 1.10 (0.03) | 0.409 (0.03) | 0.500 (0.08) | 0.404 (0.02) | 1.18 (0.05) | **0.339 (0.002)**† | 1.20 (0.05) |
| DTLZ2 d20 m3 | 1.09 (0.02) | 0.758 (0.1)† | 0.870 (0.08)† | **0.752 (0.05)**† | 1.12 (0.03) | 0.797 (0.1)† | 1.15 (0.03) |
| DTLZ2 m5 | 0.725 (0.01) | 0.559 (0.04)† | 0.556 (0.04)† | 0.642 (0.02) | 0.772 (0.02) | 0.583 (0.02) | **0.494 (0.003)**† |
| DTLZ7 m5 | 0.273 (0.005) | 0.0829 (0.006) | 0.0584 (0.007)† | **0.0454 (0.009)**† | 0.150 (0.004) | 0.0779 (0.01)† | 0.0790 (0.01) |
| Penicillin | 0.0966 (0.002) | 0.0808 (0.003)† | 0.0878 (0.003) | **0.0776 (0.002)**† | 0.0862 (0.002) | 0.103 (0.004) | 0.101 (0.002) |
| VehicleSafety | 0.215 (0.008) | 0.0715 (0.009)† | **0.0540 (0.003)**† | 0.0632 (0.006)† | 0.0685 (0.003) | 0.0904 (0.007) | 0.0945 (0.005) |
| WFG6 m4 | 0.157 (0.002) | **0.132 (0.004)**† | 0.136 (0.005)† | 0.153 (0.002) | 0.153 (0.003) | 0.150 (0.003) | 0.140 (0.004)† |
| ZDT1 d20 m2 | 2.30 (0.2) | **0.182 (0.03)**† | 0.253 (0.03)† | 0.300 (0.02) | 0.463 (0.02) | 0.562 (0.06) | 0.541 (0.09) |
| ZDT2 d20 m2 | 3.32 (0.1) | 0.136 (0.03)† | **0.134 (0.01)**† | 0.200 (0.03)† | 0.791 (0.06) | 0.832 (0.2) | 0.732 (0.1) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.144 (0.002) | **0.0669 (0.001)**† | 0.0704 (0.003)† | 0.0948 (0.005) | 0.115 (0.002) | 0.100 (0.005) | 0.111 (0.004) |
| DTLZ1 d15 m2 | 762 (50) | 361 (30)† | 355 (30)† | **300 (30)**† | 395 (50) | 384 (40)† | 404 (30) |
| DTLZ1 d15 m3 | 637 (60) | 331 (50)† | 350 (40)† | **309 (40)**† | 352 (50)† | 330 (40)† | 312 (10)† |
| DTLZ1 d20 m2 | 1105 (90) | 504 (40)† | 462 (30)† | **433 (40)**† | 478 (30)† | 520 (60)† | 453 (30)† |
| DTLZ1 d20 m3 | 917 (60) | 445 (40) | **368 (40)**† | 440 (60)† | 529 (40) | 499 (50) | 458 (40) |
| DTLZ2 d15 m2 | 0.870 (0.04) | 0.350 (0.03)† | 0.380 (0.04)† | 0.336 (0.03)† | 0.800 (0.03) | **0.335 (0.004)**† | 0.799 (0.03) |
| DTLZ2 d15 m3 | 0.785 (0.03) | 0.536 (0.06)† | 0.554 (0.05) | 0.509 (0.04)† | 0.804 (0.02) | **0.429 (0.008)**† | 0.852 (0.03) |
| DTLZ2 d20 m2 | 1.25 (0.05) | 0.442 (0.03) | 0.444 (0.05) | 0.381 (0.04)† | 1.14 (0.06) | **0.339 (0.003)**† | 1.07 (0.09) |
| DTLZ2 d20 m3 | 1.22 (0.05) | 0.865 (0.1)† | 0.891 (0.1)† | 0.854 (0.06)† | 1.18 (0.05) | **0.721 (0.1)**† | 1.20 (0.05) |
| DTLZ2 m5 | 0.744 (0.02) | 0.560 (0.05)† | 0.578 (0.04)† | 0.625 (0.04) | 0.771 (0.02) | 0.620 (0.04) | **0.488 (0.004)**† |
| DTLZ7 m5 | 0.274 (0.006) | 0.0695 (0.008) | 0.0600 (0.004) | **0.0339 (0.008)**† | 0.147 (0.004) | 0.0756 (0.01) | 0.0689 (0.006) |
| Penicillin | 0.0905 (0.002) | 0.0856 (0.002) | 0.0846 (0.002) | **0.0721 (0.003)**† | 0.0822 (0.004)† | 0.0927 (0.003) | 0.101 (0.003) |
| VehicleSafety | 0.222 (0.008) | 0.0672 (0.008)† | **0.0522 (0.003)**† | 0.0650 (0.007)† | 0.0664 (0.003) | 0.0877 (0.01) | 0.0817 (0.008) |
| WFG6 m4 | 0.151 (0.002) | **0.120 (0.004)**† | 0.125 (0.002)† | 0.150 (0.003) | 0.146 (0.004) | 0.146 (0.004) | 0.135 (0.004) |
| ZDT1 d20 m2 | 2.32 (0.2) | **0.131 (0.01)**† | 0.132 (0.006)† | 0.215 (0.03)† | 0.465 (0.03) | 0.592 (0.1) | 0.758 (0.2) |
| ZDT2 d20 m2 | 3.29 (0.09) | **0.0717 (0.01)**† | 0.124 (0.02) | 0.356 (0.05) | 0.929 (0.08) | 0.763 (0.1) | 1.10 (0.3) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.136 (0.002) | 0.0698 (0.003)† | **0.0637 (0.002)**† | 0.0903 (0.002) | 0.115 (0.001) | 0.0921 (0.004) | 0.106 (0.004) |
| DTLZ1 d15 m2 | 718 (40) | 361 (20)† | 334 (10)† | **321 (10)**† | 340 (10)† | 333 (10)† | 359 (20)† |
| DTLZ1 d15 m3 | 619 (40) | 328 (20)† | 351 (10)† | **305 (20)**† | 349 (20)† | 377 (10) | 370 (10) |
| DTLZ1 d20 m2 | 1060 (20) | 456 (20)† | 431 (10)† | **412 (8)**† | 480 (30)† | 464 (20)† | 471 (20)† |
| DTLZ1 d20 m3 | 924 (40) | 502 (20) | 468 (20) | **355 (20)**† | 525 (40) | 486 (20) | 496 (20) |
| DTLZ2 d15 m2 | 0.639 (0.02) | 0.279 (0.03)† | 0.245 (0.02)† | **0.243 (0.02)**† | 0.715 (0.04) | 0.326 (0.004) | 0.724 (0.05) |
| DTLZ2 d15 m3 | 0.664 (0.02) | 0.468 (0.06)† | 0.467 (0.05)† | 0.502 (0.03) | 0.700 (0.03) | **0.413 (0.006)**† | 0.608 (0.06) |
| DTLZ2 d20 m2 | 0.953 (0.03) | 0.616 (0.06) | 0.468 (0.07)† | 0.411 (0.04)† | 1.02 (0.03) | **0.339 (0.003)**† | 1.05 (0.02) |
| DTLZ2 d20 m3 | 0.981 (0.02) | 0.735 (0.1) | 0.899 (0.07) | 0.879 (0.05) | 0.988 (0.04) | **0.433 (0.006)**† | 1.10 (0.02) |
| DTLZ2 m5 | 0.671 (0.01) | 0.601 (0.03) | 0.633 (0.03) | 0.643 (0.03) | 0.684 (0.02) | 0.576 (0.008) | **0.497 (0.009)**† |
| DTLZ7 m5 | 0.271 (0.004) | **0.0221 (0.009)**† | 0.0886 (0.006) | 0.0425 (0.01)† | 0.157 (0.004) | 0.100 (0.02) | 0.107 (0.02) |
| Penicillin | 0.0829 (0.001) | 0.0737 (0.006)† | 0.0769 (0.001) | **0.0637 (0.003)**† | 0.0775 (0.002) | 0.0940 (0.003) | 0.0899 (0.003) |
| VehicleSafety | 0.211 (0.01) | 0.0595 (0.001)† | **0.0580 (0.002)**† | 0.0620 (0.004)† | 0.0742 (0.002) | 0.0933 (0.004) | 0.0867 (0.006) |
| WFG6 m4 | **0.152 (0.002)**† | 0.178 (0.009) | 0.180 (0.009) | 0.166 (0.003) | 0.175 (0.004) | 0.188 (0.005) | 0.200 (0.006) |
| ZDT1 d20 m2 | 2.08 (0.1) | 0.469 (0.04) | **0.251 (0.02)**† | 0.355 (0.04) | 0.515 (0.03) | 0.486 (0.04) | 0.560 (0.06) |
| ZDT2 d20 m2 | 3.36 (0.1) | 0.597 (0.02) | **0.300 (0.007)**† | 0.469 (0.04) | 0.930 (0.09) | 0.654 (0.09) | 1.08 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.150 (0.003) | 0.0881 (0.004)† | **0.0818 (0.003)**† | 0.108 (0.004) | 0.121 (0.002) | 0.0978 (0.005) | 0.126 (0.004) |
| DTLZ1 d15 m2 | 683 (40) | 283 (10)† | 261 (10)† | 265 (9)† | 303 (20) | 256 (3)† | **252 (6)**† |
| DTLZ1 d15 m3 | 545 (40) | 247 (10)† | 251 (10)† | **229 (10)**† | 286 (20)† | 233 (10)† | 247 (10)† |
| DTLZ1 d20 m2 | 1008 (50) | 379 (20)† | 347 (4)† | 346 (7)† | 395 (20)† | **345 (6)**† | 358 (10)† |
| DTLZ1 d20 m3 | 843 (40) | 332 (20)† | 330 (10)† | **306 (10)**† | 400 (30) | 332 (9)† | 352 (20)† |
| DTLZ2 d15 m2 | 0.743 (0.04) | 0.222 (0.03)† | **0.193 (0.009)**† | 0.264 (0.03) | 0.804 (0.03) | 0.332 (0.003) | 0.837 (0.03) |
| DTLZ2 d15 m3 | 0.771 (0.02) | 0.501 (0.07)† | **0.360 (0.03)**† | 0.477 (0.06)† | 0.741 (0.03) | 0.418 (0.004)† | 0.669 (0.07) |
| DTLZ2 d20 m2 | 1.07 (0.03) | 0.442 (0.1)† | 0.551 (0.1)† | 0.407 (0.05)† | 1.14 (0.05) | **0.331 (0.003)**† | 1.19 (0.04) |
| DTLZ2 d20 m3 | 1.09 (0.03) | 0.813 (0.1) | 0.639 (0.1)† | 0.734 (0.08) | 1.04 (0.04) | **0.417 (0.003)**† | 1.14 (0.02) |
| DTLZ2 m5 | 0.749 (0.02) | 0.694 (0.05)† | 0.612 (0.06)† | 0.735 (0.04) | 0.735 (0.03) | 0.678 (0.03) | **0.496 (0.01)**† |
| DTLZ7 m5 | 0.272 (0.005) | 0.0720 (0.01)† | 0.0624 (0.005)† | **0.0508 (0.01)**† | 0.146 (0.005) | 0.166 (0.02) | 0.0983 (0.01) |
| Penicillin | 0.0898 (0.001) | 0.0799 (0.003)† | 0.0819 (0.003) | **0.0736 (0.002)**† | 0.0852 (0.002) | 0.105 (0.004) | 0.107 (0.004) |
| VehicleSafety | 0.214 (0.01) | 0.0623 (0.005) | 0.0569 (0.003)† | **0.0501 (0.004)**† | 0.0753 (0.005) | 0.0900 (0.009) | 0.0695 (0.005) |
| WFG6 m4 | 0.156 (0.002) | 0.147 (0.004)† | **0.139 (0.005)**† | 0.159 (0.002) | 0.160 (0.003) | 0.157 (0.002) | 0.142 (0.003)† |
| ZDT1 d20 m2 | 2.04 (0.09) | 0.150 (0.01) | **0.0882 (0.006)**† | 0.230 (0.02) | 0.339 (0.01) | 0.245 (0.05) | 0.358 (0.05) |
| ZDT2 d20 m2 | 3.29 (0.1) | 0.0910 (0.02)† | **0.0642 (0.009)**† | 0.186 (0.05)† | 0.973 (0.07) | 0.372 (0.04) | 0.769 (0.1) |

Table 19: Terminal nIGD$^+$ at $\eta = 0.2$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.282 (0.007) | 0.236 (0.008)† | **0.219 (0.009)**† | 0.236 (0.02)† | 0.246 (0.008)† | 0.221 (0.01)† | 0.222 (0.006)† |
| DTLZ1 d15 m2 | 920 (60) | 484 (90)† | 489 (60)† | 488 (60)† | 682 (109)† | 490 (30)† | **410 (30)**† |
| DTLZ1 d15 m3 | 819 (50) | 527 (40)† | 598 (60)† | 668 (60) | 639 (60) | 540 (80)† | **485 (80)**† |
| DTLZ1 d20 m2 | 1339 (70) | 665 (109)† | 604 (40)† | 808 (70)† | **595 (70)**† | 616 (50)† | 617 (50)† |
| DTLZ1 d20 m3 | 1207 (60) | 817 (103) | 910 (103) | 851 (70) | 792 (127)† | 658 (40)† | **556 (40)**† |
| DTLZ2 d15 m2 | 1.11 (0.05) | 1.09 (0.04) | 0.967 (0.07) | 1.04 (0.07) | 0.985 (0.03) | **0.669 (0.1)**† | 0.797 (0.1)† |
| DTLZ2 d15 m3 | 1.11 (0.02)† | 1.21 (0.06) | 1.17 (0.06)† | 1.01 (0.05)† | 0.922 (0.05)† | 0.961 (0.08)† | **0.822 (0.1)**† |
| DTLZ2 d20 m2 | 1.47 (0.06) | 1.49 (0.08) | 1.49 (0.08) | 1.38 (0.1) | 1.30 (0.05) | **0.509 (0.1)**† | 1.30 (0.1) |
| DTLZ2 d20 m3 | 1.47 (0.04) | 1.54 (0.05) | 1.57 (0.06) | 1.41 (0.06) | **1.21 (0.07)**† | 1.34 (0.1)† | 1.33 (0.1)† |
| DTLZ2 m5 | 1.04 (0.04) | 1.03 (0.05) | 1.09 (0.03) | 0.956 (0.04)† | 0.922 (0.04)† | 1.03 (0.07) | **0.755 (0.09)**† |
| DTLZ7 m5 | 0.422 (0.01) | 0.167 (0.01) | 0.147 (0.01) | 0.196 (0.01) | 0.222 (0.01) | 0.106 (0.01)† | **0.0931 (0.01)**† |
| Penicillin | 0.179 (0.007) | **0.133 (0.005)**† | 0.156 (0.009) | 0.160 (0.01)† | 0.150 (0.005) | 0.159 (0.003) | 0.157 (0.01) |
| VehicleSafety | 0.394 (0.02) | 0.143 (0.01)† | 0.175 (0.02)† | 0.151 (0.01)† | 0.168 (0.02)† | 0.156 (0.02)† | **0.122 (0.02)**† |
| WFG6 m4 | 0.291 (0.004) | 0.266 (0.004) | 0.272 (0.005) | 0.262 (0.003) | 0.253 (0.005) | **0.230 (0.007)**† | 0.237 (0.01)† |
| ZDT1 d20 m2 | 2.87 (0.1) | **1.60 (0.2)**† | **1.60 (0.2)**† | 1.82 (0.2)† | 1.96 (0.2)† | 1.89 (0.2)† | 1.96 (0.2)† |
| ZDT2 d20 m2 | 4.04 (0.2) | **1.28 (0.2)**† | 1.49 (0.2)† | 2.09 (0.3) | 1.54 (0.2)† | 2.47 (0.3) | 2.31 (0.3) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.213 (0.004) | 0.178 (0.007)† | **0.171 (0.008)**† | 0.178 (0.006)† | 0.194 (0.006) | 0.176 (0.01)† | 0.206 (0.01) |
| DTLZ1 d15 m2 | 1030 (90) | 463 (40)† | 524 (80)† | 648 (133)† | 621 (136)† | 414 (30)† | **364 (20)**† |
| DTLZ1 d15 m3 | 766 (90) | 432 (60)† | 390 (50)† | 448 (50)† | 428 (30)† | **374 (30)**† | 377 (30)† |
| DTLZ1 d20 m2 | 1337 (128) | 787 (160)† | 592 (50)† | 832 (144)† | 852 (201)† | 631 (30)† | **557 (60)**† |
| DTLZ1 d20 m3 | 1158 (106) | 642 (80)† | 716 (90)† | 651 (70)† | **590 (90)**† | 616 (50)† | 592 (60)† |
| DTLZ2 d15 m2 | 1.17 (0.04) | 1.30 (0.09) | 1.19 (0.2) | 1.27 (0.1) | 1.05 (0.07) | **0.647 (0.2)**† | 1.04 (0.2)† |
| DTLZ2 d15 m3 | 1.01 (0.03)† | 1.21 (0.09)† | 1.20 (0.07)† | 1.10 (0.1)† | 1.09 (0.2)† | **0.935 (0.2)**† | 0.976 (0.1)† |
| DTLZ2 d20 m2 | 1.66 (0.06) | 1.51 (0.09) | 1.63 (0.3) | 1.37 (0.2) | 1.56 (0.2) | **0.651 (0.2)**† | 1.73 (0.3) |
| DTLZ2 d20 m3 | 1.46 (0.05) | 1.69 (0.1) | 1.91 (0.1) | **1.28 (0.2)**† | 1.35 (0.1)† | 1.51 (0.2)† | 1.53 (0.07) |
| DTLZ2 m5 | 0.935 (0.02)† | 0.993 (0.03) | 1.03 (0.07)† | 0.853 (0.04)† | 0.844 (0.03)† | 0.951 (0.06)† | **0.800 (0.1)**† |
| DTLZ7 m5 | 0.356 (0.007) | 0.124 (0.01) | 0.112 (0.008) | 0.151 (0.02) | 0.179 (0.005) | **0.0818 (0.01)**† | 0.0882 (0.01)† |
| Penicillin | 0.142 (0.006) | 0.129 (0.005) | 0.116 (0.007)† | **0.108 (0.005)**† | 0.139 (0.009) | 0.152 (0.01) | 0.155 (0.02) |
| VehicleSafety | 0.307 (0.01) | 0.114 (0.01)† | 0.124 (0.01)† | 0.123 (0.02)† | 0.173 (0.02) | 0.111 (0.02)† | **0.110 (0.009)**† |
| WFG6 m4 | 0.220 (0.004) | 0.219 (0.004) | 0.204 (0.004)† | 0.208 (0.004)† | 0.202 (0.004)† | **0.192 (0.007)**† | 0.201 (0.01)† |
| ZDT1 d20 m2 | 3.74 (0.1) | 1.36 (0.3)† | **1.05 (0.1)**† | 1.82 (0.3) | 2.21 (0.4)† | 2.07 (0.3) | 2.91 (0.3) |
| ZDT2 d20 m2 | 4.63 (0.1) | 1.25 (0.4)† | **0.944 (0.3)**† | 1.87 (0.3) | 1.53 (0.3) | 3.05 (0.3) | 3.14 (0.3) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.176 (0.003) | 0.158 (0.01)† | **0.139 (0.01)**† | 0.145 (0.01)† | 0.162 (0.004)† | 0.165 (0.008) | 0.185 (0.02) |
| DTLZ1 d15 m2 | 874 (30) | 658 (60) | 619 (50) | 502 (30)† | 647 (70) | **427 (30)**† | 475 (40)† |
| DTLZ1 d15 m3 | 877 (90) | 594 (70)† | 586 (50)† | 666 (50) | 637 (60) | **459 (40)**† | 536 (40)† |
| DTLZ1 d20 m2 | 1137 (60) | 857 (90)† | 739 (60)† | 787 (50)† | 1075 (40) | **630 (40)**† | 642 (40)† |
| DTLZ1 d20 m3 | 1333 (113) | 1042 (125) | 925 (70) | 1127 (110) | 1247 (101) | 613 (50)† | **570 (60)**† |
| DTLZ2 d15 m2 | 0.693 (0.02) | 0.671 (0.03) | 0.669 (0.04) | 0.608 (0.04)† | 0.629 (0.02) | **0.538 (0.03)**† | 0.667 (0.07)† |
| DTLZ2 d15 m3 | 0.776 (0.01) | 0.822 (0.03) | 0.799 (0.02) | 0.729 (0.02)† | **0.689 (0.02)**† | 0.837 (0.03) | 0.814 (0.04)† |
| DTLZ2 d20 m2 | 1.00 (0.02) | 1.01 (0.04) | 0.971 (0.04) | 0.848 (0.04) | 0.776 (0.02) | **0.542 (0.06)**† | 1.03 (0.07) |
| DTLZ2 d20 m3 | 1.06 (0.01) | 1.13 (0.03) | 1.08 (0.04) | 0.935 (0.03)† | 0.840 (0.01)† | **0.813 (0.07)**† | 0.932 (0.08)† |
| DTLZ2 m5 | 0.866 (0.02) | 0.897 (0.02) | 0.866 (0.01) | **0.749 (0.03)**† | 0.783 (0.02)† | 0.829 (0.02)† | 0.911 (0.02) |
| DTLZ7 m5 | 0.295 (0.004) | **0.0188 (0.009)**† | 0.0549 (0.01) | 0.0209 (0.007)† | 0.189 (0.006) | 0.0581 (0.02)† | 0.0489 (0.01)† |
| Penicillin | 0.0953 (0.003) | **0.0784 (0.003)**† | 0.0832 (0.002)† | 0.0785 (0.003)† | 0.0841 (0.002)† | 0.100 (0.004) | 0.104 (0.003) |
| VehicleSafety | 0.318 (0.02) | 0.126 (0.007)† | 0.129 (0.01)† | **0.111 (0.009)**† | 0.134 (0.01)† | 0.123 (0.02)† | 0.120 (0.01)† |
| WFG6 m4 | **0.235 (0.009)**† | 0.313 (0.02) | 0.307 (0.02) | 0.272 (0.02)† | 0.327 (0.03) | 0.402 (0.03) | 0.415 (0.02) |
| ZDT1 d20 m2 | 2.46 (0.1) | 1.43 (0.1)† | **1.35 (0.1)**† | 2.01 (0.1) | 1.63 (0.1)† | 1.54 (0.1)† | 1.63 (0.2)† |
| ZDT2 d20 m2 | 3.57 (0.1) | 1.92 (0.2)† | 1.73 (0.3)† | **1.57 (0.1)**† | 1.62 (0.1)† | 1.68 (0.2)† | 1.82 (0.1)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.245 (0.005)† | 0.226 (0.009)† | 0.224 (0.01)† | **0.218 (0.01)**† | 0.227 (0.01)† | 0.232 (0.01)† | 0.225 (0.02)† |
| DTLZ1 d15 m2 | 901 (80) | 460 (50)† | 484 (50)† | 438 (40)† | 845 (90) | 372 (20)† | **363 (30)**† |
| DTLZ1 d15 m3 | 712 (80) | 536 (90)† | 572 (90) | 782 (80) | 788 (90) | **341 (30)**† | 349 (40)† |
| DTLZ1 d20 m2 | 1310 (90) | 636 (80)† | 694 (60) | 649 (90)† | 1421 (100) | **442 (30)**† | 476 (30)† |
| DTLZ1 d20 m3 | 1109 (90) | 597 (50)† | 923 (90) | 1150 (114) | 1244 (109) | 526 (40)† | **514 (50)**† |
| DTLZ2 d15 m2 | 1.04 (0.03) | 1.22 (0.06) | 1.10 (0.06) | 0.996 (0.06) | 0.953 (0.05) | **0.643 (0.08)**† | 0.825 (0.1)† |
| DTLZ2 d15 m3 | 1.05 (0.02) | 1.09 (0.09) | 1.14 (0.07) | 0.953 (0.06) | 0.859 (0.02) | 1.10 (0.1) | **0.708 (0.09)**† |
| DTLZ2 d20 m2 | 1.38 (0.05) | 1.46 (0.06) | 1.43 (0.05) | 1.45 (0.06) | 1.20 (0.08)† | **0.994 (0.2)**† | 1.44 (0.1) |
| DTLZ2 d20 m3 | 1.44 (0.02) | 1.64 (0.06) | 1.61 (0.04) | 1.25 (0.06) | **1.00 (0.04)**† | 1.61 (0.07) | 1.20 (0.2)† |
| DTLZ2 m5 | 0.989 (0.02) | 1.09 (0.04) | 1.10 (0.02) | 0.972 (0.04)† | **0.860 (0.03)**† | 1.33 (0.09) | 1.06 (0.2)† |
| DTLZ7 m5 | 0.372 (0.008) | 0.135 (0.02)† | 0.146 (0.007)† | **0.115 (0.01)**† | 0.168 (0.009) | 0.134 (0.05)† | 0.147 (0.02)† |
| Penicillin | 0.166 (0.008) | 0.132 (0.007)† | **0.130 (0.006)**† | 0.145 (0.005)† | 0.134 (0.01)† | 0.150 (0.007)† | 0.147 (0.008)† |
| VehicleSafety | 0.364 (0.02) | 0.184 (0.02) | 0.191 (0.02) | 0.141 (0.01)† | 0.216 (0.01) | 0.129 (0.01)† | **0.113 (0.01)**† |
| WFG6 m4 | 0.257 (0.004)† | 0.265 (0.004) | 0.256 (0.003)† | 0.243 (0.004)† | 0.255 (0.008)† | **0.224 (0.02)**† | 0.248 (0.009)† |
| ZDT1 d20 m2 | 2.69 (0.2) | 1.14 (0.2) | 1.07 (0.2) | 1.64 (0.2) | 2.55 (0.2) | **0.578 (0.06)**† | 1.01 (0.2) |
| ZDT2 d20 m2 | 3.74 (0.2) | 0.745 (0.1)† | **0.621 (0.1)**† | 0.720 (0.1)† | 2.22 (0.3) | 1.15 (0.2) | 1.41 (0.2) |

Table 20: Terminal nIGD$^+$ at $\eta = 0.4$, likelihood-free estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.335 (0.009)† | **0.313 (0.02)**† | 0.320 (0.01)† | 0.343 (0.008)† | 0.336 (0.01)† | 0.346 (0.02)† | 0.337 (0.02)† |
| DTLZ1 d15 m2 | 1030 (90) | 811 (60) | 986 (137) | 925 (100) | 1065 (90) | 607 (110)† | **544 (70)**† |
| DTLZ1 d15 m3 | 1106 (100) | 849 (100) | 867 (60) | 853 (70) | 765 (80)† | 587 (80)† | **585 (80)**† |
| DTLZ1 d20 m2 | 1467 (90) | 1033 (104) | 1317 (137) | 1297 (129) | 1593 (119) | **693 (80)**† | 699 (60)† |
| DTLZ1 d20 m3 | 1325 (80) | 1136 (90) | 1300 (60) | 1324 (70) | 1516 (90) | **731 (80)**† | 867 (126)† |
| DTLZ2 d15 m2 | 1.27 (0.09)† | 1.21 (0.09)† | 1.33 (0.07)† | 1.13 (0.1)† | 1.19 (0.09)† | 0.963 (0.2)† | **0.895 (0.2)**† |
| DTLZ2 d15 m3 | 1.32 (0.04)† | 1.40 (0.08)† | 1.29 (0.07)† | 1.22 (0.07)† | **1.13 (0.09)**† | 1.46 (0.2)† | 1.23 (0.2)† |
| DTLZ2 d20 m2 | 1.55 (0.06)† | 1.68 (0.2)† | 1.54 (0.07)† | 1.62 (0.08)† | 1.56 (0.2)† | **1.29 (0.2)**† | 1.56 (0.1)† |
| DTLZ2 d20 m3 | 1.66 (0.07)† | 1.57 (0.1)† | **1.48 (0.09)**† | 1.50 (0.09)† | 1.73 (0.2)† | 1.76 (0.07) | 1.79 (0.05) |
| DTLZ2 m5 | 1.25 (0.05) | 1.13 (0.06)† | 1.16 (0.07)† | **1.05 (0.04)**† | 1.14 (0.07)† | 1.62 (0.2) | 1.32 (0.07) |
| DTLZ7 m5 | 0.598 (0.01) | 0.332 (0.03)† | 0.353 (0.02)† | 0.403 (0.03) | 0.435 (0.02) | 0.232 (0.03)† | **0.217 (0.04)**† |
| Penicillin | 0.239 (0.02)† | **0.216 (0.01)**† | 0.223 (0.02)† | 0.235 (0.02)† | 0.224 (0.02)† | 0.243 (0.02)† | 0.235 (0.02)† |
| VehicleSafety | 0.610 (0.03) | 0.296 (0.03) | 0.293 (0.03) | 0.386 (0.03) | 0.343 (0.03) | **0.182 (0.02)**† | 0.239 (0.06)† |
| WFG6 m4 | 0.387 (0.009) | 0.378 (0.01) | 0.378 (0.01) | 0.356 (0.01)† | 0.368 (0.006) | **0.328 (0.01)**† | 0.334 (0.02)† |
| ZDT1 d20 m2 | 3.16 (0.1)† | 2.83 (0.2)† | 3.11 (0.3)† | 3.16 (0.1)† | 3.37 (0.2)† | 3.20 (0.2)† | **2.62 (0.3)**† |
| ZDT2 d20 m2 | 4.47 (0.2) | 3.17 (0.2)† | **3.07 (0.3)**† | 3.95 (0.2) | 3.83 (0.3)† | 3.15 (0.3)† | 3.14 (0.4)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.283 (0.007)† | 0.283 (0.02)† | 0.280 (0.01)† | 0.280 (0.009)† | 0.290 (0.01) | **0.250 (0.01)**† | 0.274 (0.02)† |
| DTLZ1 d15 m2 | 1043 (50) | 665 (102)† | 687 (100)† | 735 (90)† | 715 (123)† | 557 (60)† | **464 (40)**† |
| DTLZ1 d15 m3 | 796 (60) | 592 (105)† | 705 (50) | 646 (60) | 632 (90) | 585 (70) | **382 (30)**† |
| DTLZ1 d20 m2 | 1337 (109) | 960 (123)† | 1007 (138)† | 1067 (80) | 817 (153)† | 755 (70)† | **655 (80)**† |
| DTLZ1 d20 m3 | 1361 (70) | 1108 (107) | 1140 (107) | 1148 (104) | 883 (143)† | **726 (125)**† | 766 (132)† |
| DTLZ2 d15 m2 | **1.13 (0.04)**† | 1.23 (0.1)† | 1.23 (0.1)† | 1.19 (0.1)† | 1.50 (0.07) | 1.64 (0.4)† | 1.93 (0.3) |
| DTLZ2 d15 m3 | 1.14 (0.04)† | 1.42 (0.1)† | 1.14 (0.07)† | 1.29 (0.06)† | 1.34 (0.3)† | 1.55 (0.2)† | **1.06 (0.2)**† |
| DTLZ2 d20 m2 | **1.51 (0.09)**† | 1.74 (0.2)† | 1.58 (0.1)† | 1.63 (0.1)† | 1.69 (0.2)† | 1.69 (0.3)† | 2.16 (0.5)† |
| DTLZ2 d20 m3 | 1.58 (0.05) | 1.41 (0.05)† | 1.43 (0.2)† | **1.36 (0.09)**† | 1.55 (0.2)† | 2.07 (0.4) | 2.41 (0.4) |
| DTLZ2 m5 | 1.14 (0.03)† | 1.18 (0.08)† | 1.13 (0.09)† | 1.06 (0.05)† | **1.05 (0.05)**† | 1.36 (0.2) | 1.43 (0.2) |
| DTLZ7 m5 | 0.461 (0.01) | 0.276 (0.02) | 0.283 (0.02) | 0.301 (0.02) | 0.297 (0.02) | 0.165 (0.02)† | **0.157 (0.03)**† |
| Penicillin | 0.185 (0.02)† | **0.161 (0.007)**† | 0.173 (0.02)† | 0.168 (0.01)† | 0.197 (0.01) | 0.243 (0.02) | 0.224 (0.02) |
| VehicleSafety | 0.445 (0.03) | 0.200 (0.05)† | 0.193 (0.02)† | 0.269 (0.03) | 0.210 (0.03)† | **0.158 (0.02)**† | 0.161 (0.03)† |
| WFG6 m4 | 0.298 (0.007) | 0.303 (0.006) | 0.294 (0.006) | 0.313 (0.008) | 0.284 (0.01)† | **0.255 (0.01)**† | **0.255 (0.01)**† |
| ZDT1 d20 m2 | 3.87 (0.3) | 2.77 (0.4) | **1.77 (0.2)**† | 2.56 (0.09) | 3.38 (0.4) | 2.99 (0.3) | 3.02 (0.4) |
| ZDT2 d20 m2 | 4.70 (0.2) | 3.01 (0.4)† | **2.44 (0.3)**† | 3.30 (0.3)† | 3.05 (0.4)† | 3.77 (0.2) | 4.20 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.202 (0.004)† | 0.222 (0.009) | **0.195 (0.008)**† | 0.202 (0.007)† | 0.199 (0.007)† | 0.205 (0.007)† | 0.239 (0.02)† |
| DTLZ1 d15 m2 | 1118 (60) | 784 (111)† | 909 (90) | 991 (90) | 1100 (60) | 698 (100)† | **563 (100)**† |
| DTLZ1 d15 m3 | 1077 (80) | 1004 (60) | 961 (112) | 1031 (70) | 1055 (60) | 882 (80)† | **657 (100)**† |
| DTLZ1 d20 m2 | 1481 (70) | 1302 (134) | 1184 (101) | 1101 (130)† | 1590 (90) | **894 (80)**† | 917 (100)† |
| DTLZ1 d20 m3 | 1459 (112)† | 1479 (80)† | 1426 (127)† | 1389 (90)† | 1590 (70) | 1302 (149)† | **1163 (163)**† |
| DTLZ2 d15 m2 | 0.984 (0.03) | 0.873 (0.05) | 0.908 (0.05) | 0.924 (0.07) | 1.00 (0.03) | **0.356 (0.001)**† | 0.358 (0.001)† |
| DTLZ2 d15 m3 | 1.18 (0.04) | 0.944 (0.06) | 0.981 (0.04) | 0.922 (0.08) | 1.09 (0.04) | 0.777 (0.1) | **0.433 (0.002)**† |
| DTLZ2 d20 m2 | 1.29 (0.03) | 1.09 (0.05) | 1.16 (0.05) | 1.04 (0.05) | 1.22 (0.04) | **0.358 (0.002)**† | 0.359 (0.002)† |
| DTLZ2 d20 m3 | 1.48 (0.04) | 1.44 (0.1) | 1.24 (0.09) | 1.18 (0.1) | 1.39 (0.09) | 1.26 (0.1) | **0.464 (0.01)**† |
| DTLZ2 m5 | 1.11 (0.03)† | **1.02 (0.04)**† | 1.08 (0.04)† | 1.07 (0.03)† | 1.22 (0.09)† | 1.78 (0.03) | 1.08 (0.1)† |
| DTLZ7 m5 | 0.319 (0.005) | 0.0967 (0.03)† | **0.0435 (0.01)**† | 0.0735 (0.03)† | 0.232 (0.01) | 0.0593 (0.02)† | 0.0990 (0.03)† |
| Penicillin | 0.102 (0.004)† | 0.0955 (0.004)† | **0.0925 (0.003)**† | 0.0945 (0.004)† | 0.0996 (0.002)† | 0.102 (0.003)† | 0.0986 (0.004)† |
| VehicleSafety | 0.452 (0.02) | **0.119 (0.02)**† | 0.267 (0.03) | 0.193 (0.02) | 0.261 (0.03) | 0.143 (0.02)† | 0.174 (0.01) |
| WFG6 m4 | **0.357 (0.02)**† | 0.430 (0.03)† | 0.415 (0.02)† | 0.370 (0.03)† | 0.478 (0.03) | 0.477 (0.01) | 0.442 (0.02) |
| ZDT1 d20 m2 | 3.42 (0.2)† | 3.05 (0.2)† | 3.19 (0.2)† | 3.32 (0.2)† | 3.65 (0.3)† | **2.90 (0.1)**† | 3.07 (0.2)† |
| ZDT2 d20 m2 | 4.07 (0.1) | 2.94 (0.2)† | 3.30 (0.2)† | 3.07 (0.3)† | 3.45 (0.3)† | **2.76 (0.2)**† | 3.10 (0.2)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.331 (0.01) | **0.276 (0.008)**† | 0.319 (0.009) | 0.292 (0.01)† | 0.309 (0.01) | 0.314 (0.03)† | 0.317 (0.02)† |
| DTLZ1 d15 m2 | 1067 (60) | 877 (90) | 759 (119) | 1075 (90) | 1094 (80) | **392 (20)**† | 414 (30)† |
| DTLZ1 d15 m3 | 985 (105) | 689 (120) | 787 (104) | 819 (70) | 984 (60) | 435 (40)† | **356 (30)**† |
| DTLZ1 d20 m2 | 1567 (129) | 1164 (100) | 1329 (123) | 1444 (143) | 1629 (116) | 573 (50)† | **500 (20)**† |
| DTLZ1 d20 m3 | 1181 (111) | 1245 (90) | 1151 (115) | 1282 (100) | 1436 (80) | 581 (60)† | **517 (30)**† |
| DTLZ2 d15 m2 | 1.13 (0.04)† | 1.15 (0.07)† | 1.19 (0.07)† | 1.08 (0.09)† | **1.05 (0.06)**† | 1.82 (0.1) | 1.44 (0.2)† |
| DTLZ2 d15 m3 | 1.19 (0.02) | 1.09 (0.05)† | 1.12 (0.06)† | **1.02 (0.05)**† | 1.04 (0.03)† | 1.71 (0.1) | 1.55 (0.1) |
| DTLZ2 d20 m2 | 1.59 (0.05)† | 1.41 (0.06)† | 1.48 (0.09)† | **1.33 (0.09)**† | 1.44 (0.03)† | 2.24 (0.1) | 1.46 (0.2)† |
| DTLZ2 d20 m3 | 1.55 (0.05) | 1.55 (0.06) | 1.60 (0.05) | **1.34 (0.04)**† | 1.43 (0.07)† | 2.27 (0.1) | 2.05 (0.2) |
| DTLZ2 m5 | 1.13 (0.02) | 1.09 (0.03) | 1.17 (0.04) | **0.972 (0.04)**† | 1.04 (0.04)† | 1.78 (0.1) | 1.72 (0.08) |
| DTLZ7 m5 | 0.508 (0.008) | 0.239 (0.01) | 0.284 (0.02) | 0.336 (0.01) | 0.281 (0.02) | **0.117 (0.03)**† | 0.261 (0.05) |
| Penicillin | 0.232 (0.02) | 0.175 (0.01)† | **0.166 (0.01)**† | 0.203 (0.02)† | 0.193 (0.01)† | 0.198 (0.009)† | 0.227 (0.02) |
| VehicleSafety | 0.542 (0.03) | 0.299 (0.03) | 0.283 (0.02) | 0.248 (0.03)† | 0.433 (0.03) | 0.190 (0.02)† | **0.167 (0.02)**† |
| WFG6 m4 | 0.326 (0.007) | 0.347 (0.01) | 0.328 (0.007) | 0.315 (0.009) | 0.310 (0.01) | **0.211 (0.01)**† | 0.228 (0.02)† |
| ZDT1 d20 m2 | 2.81 (0.2) | 2.50 (0.3)† | 2.45 (0.2) | 3.12 (0.3) | 2.80 (0.2) | **1.67 (0.2)**† | 2.01 (0.3)† |
| ZDT2 d20 m2 | 3.92 (0.2) | 3.39 (0.4)† | 3.18 (0.4)† | 3.84 (0.3)† | 3.99 (0.2) | **2.39 (0.5)**† | 3.06 (0.4)† |

Table 21: Terminal $\log_{10}$ HV gap at $\eta = 0.05$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.61 (0.01) | **1.56 (0.01)**† | 1.98 (0.03) | 2.00 (0.008) | 1.96 (0.03) | 1.94 (0.01) |
| DTLZ1 d15 m2 | 5.15 (0.01) | 4.69 (0.05) | 4.69 (0.06) | 4.68 (0.05) | 4.61 (0.05)† | **4.50 (0.04)**† | 4.59 (0.07)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.39 (0.02)† | 6.42 (0.06)† | 6.60 (0.04) | 6.72 (0.06) | 6.37 (0.03)† | **6.35 (0.04)**† |
| DTLZ1 d20 m2 | 5.20 (4e-4) | 4.82 (0.04)† | 4.81 (0.04)† | 4.83 (0.04)† | 4.85 (0.03)† | **4.79 (0.03)**† | 4.85 (0.05)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.84 (0.04)† | 6.91 (0.02) | 7.07 (0.04) | 7.14 (0.03) | 6.79 (0.02)† | **6.77 (0.02)**† |
| DTLZ2 d15 m2 | -0.372 (0) | **-0.661 (0.02)**† | -0.649 (0.02)† | -0.549 (0.02) | -0.372 (0) | -0.588 (0.005) | -0.423 (0.03) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.178 (0.03)† | **-0.194 (0.02)**† | -0.124 (0.01) | -0.0932 (2e-4) | -0.182 (0.006)† | -0.103 (0.01) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.511 (0.02) | -0.540 (0.03)† | -0.440 (0.02) | -0.372 (0) | **-0.574 (0.002)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | **-0.151 (0.02)**† | -0.122 (0.01)† | -0.100 (0.003)† | -0.0929 (0) | -0.128 (0.01)† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003) | 0.0452 (0.02)† | **0.0373 (0.02)**† | 0.127 (0.009) | 0.154 (0.002) | 0.124 (0.003) | 0.110 (0.002) |
| Penicillin | 5.71 (0.01) | 5.64 (0.02)† | 5.63 (0.02)† | 5.71 (0.02) | 5.64 (0.02)† | 5.61 (0.02)† | **5.60 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | 0.0951 (0.06) | **-0.366 (0.04)**† | 0.392 (0.03) | 0.280 (0.04) | 0.409 (0.06) | 0.416 (0.03) |
| WFG6 m4 | 2.65 (0.003) | **2.13 (0.009)**† | 2.15 (0.02)† | 2.40 (0.02) | 2.33 (0.03) | 2.21 (0.007) | 2.16 (0.01)† |
| ZDT1 d20 m2 | 1.41 (0.02) | **-0.0230 (0.2)**† | 0.325 (0.09)† | 0.500 (0.08) | 0.585 (0.05) | 0.742 (0.06) | 0.419 (0.2)† |
| ZDT2 d20 m2 | 1.63 (0.01) | **-0.673 (0.2)**† | -0.626 (0.1)† | -0.164 (0.4)† | 1.07 (0.05) | 1.06 (0.1) | 0.907 (0.1) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.61 (0.01)† | **1.57 (0.02)**† | 1.92 (0.03) | 2.02 (0.004) | 1.95 (0.01) | 1.94 (0.02) |
| DTLZ1 d15 m2 | 5.15 (0.01) | 4.68 (0.05)† | 4.64 (0.03)† | 4.60 (0.09)† | 4.67 (0.03)† | **4.53 (0.06)**† | 4.59 (0.06)† |
| DTLZ1 d15 m3 | 7.51 (0.02) | 6.39 (0.03)† | 6.44 (0.04)† | 6.70 (0.07) | 6.65 (0.06) | **6.38 (0.04)**† | 6.39 (0.05)† |
| DTLZ1 d20 m2 | 5.20 (0.006) | 4.81 (0.05)† | 4.85 (0.03) | 4.86 (0.03) | 4.80 (0.03)† | **4.73 (0.03)**† | 4.77 (0.03)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | **6.82 (0.04)**† | **6.82 (0.04)**† | 7.04 (0.04) | 7.17 (0.04) | 6.83 (0.02)† | 6.88 (0.06)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.600 (0.03)† | **-0.650 (0.03)**† | -0.510 (0.04) | -0.372 (0) | -0.581 (0.006) | -0.372 (0) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | **-0.197 (0.03)**† | -0.192 (0.03)† | -0.148 (0.01)† | -0.0935 (3e-4) | -0.183 (0.003)† | -0.0930 (3e-5) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.506 (0.02) | **-0.570 (0.03)**† | -0.470 (0.02) | -0.372 (0) | **-0.570 (0.007)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.129 (0.02)† | -0.132 (0.02)† | -0.107 (0.007)† | -0.0929 (0) | **-0.145 (0.01)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003) | **0.0367 (0.03)**† | 0.0418 (0.02)† | 0.124 (0.01) | 0.155 (0.003) | 0.128 (0.007) | 0.107 (0.003)† |
| Penicillin | 5.71 (0.01) | 5.69 (0.03) | 5.62 (0.02)† | 5.74 (0.02) | 5.72 (0.03) | 5.65 (0.02)† | **5.61 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | -0.0452 (0.03) | **-0.339 (0.02)**† | 0.367 (0.03) | 0.299 (0.03) | 0.484 (0.06) | 0.512 (0.06) |
| WFG6 m4 | 2.65 (0.003) | 2.13 (0.01)† | **2.12 (0.01)**† | 2.43 (0.02) | 2.34 (0.02) | 2.20 (0.005) | 2.16 (0.008) |
| ZDT1 d20 m2 | 1.41 (0.02) | 0.211 (0.08)† | **0.0895 (0.2)**† | 0.461 (0.06)† | 0.704 (0.03) | 0.496 (0.2) | 0.676 (0.1) |
| ZDT2 d20 m2 | 1.63 (0.01) | **-1.02 (0.06)**† | -0.743 (0.06) | 0.674 (0.2) | 1.14 (0.03) | 0.940 (0.2) | 0.922 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.28 (0.009) | 1.80 (0.008) | **1.75 (0.008)**† | 1.90 (0.01) | 2.07 (0.008) | 2.02 (0.01) | 1.96 (0.02) |
| DTLZ1 d15 m2 | 5.12 (0.02) | 4.73 (0.05) | 4.78 (0.05) | 4.71 (0.04) | 4.77 (0.07) | **4.57 (0.03)**† | 4.66 (0.04)† |
| DTLZ1 d15 m3 | 7.52 (0.02) | 6.80 (0.09) | 6.67 (0.04) | 6.72 (0.05) | 6.84 (0.05) | **6.55 (0.04)**† | 6.66 (0.05)† |
| DTLZ1 d20 m2 | 5.20 (0) | 4.91 (0.03)† | 4.94 (0.04)† | 4.92 (0.05)† | 5.00 (0.05) | **4.85 (0.04)**† | 4.92 (0.04)† |
| DTLZ1 d20 m3 | 7.76 (0.007) | 7.13 (0.05) | 6.99 (0.04)† | 7.08 (0.04) | 7.40 (0.06) | **6.94 (0.02)**† | 7.14 (0.04) |
| DTLZ2 d15 m2 | -0.372 (0) | -0.556 (0.03) | **-0.628 (0.02)**† | -0.510 (0.03) | -0.372 (0) | -0.551 (0.02) | -0.394 (0.02) |
| DTLZ2 d15 m3 | -0.0966 (0.002) | -0.197 (0.02)† | **-0.200 (0.02)**† | -0.134 (0.01)† | -0.0973 (0.004) | -0.191 (0.004)† | -0.136 (0.01) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.401 (0.02) | -0.484 (0.03)† | -0.431 (0.02) | -0.372 (0) | **-0.557 (0.005)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.138 (0.02)† | -0.109 (0.009) | -0.0970 (0.003) | -0.0932 (2e-4) | **-0.180 (0.002)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.140 (0.002)† | **0.0895 (0.02)**† | 0.102 (0.02)† | 0.145 (0.005) | 0.153 (0.003) | 0.144 (0.007) | 0.0929 (0.006)† |
| Penicillin | 5.70 (0.01) | 5.65 (0.05)† | 5.62 (0.03)† | 5.72 (0.05) | 5.67 (0.02) | 5.71 (0.04) | **5.59 (0.02)**† |
| VehicleSafety | 1.24 (0.01) | 0.870 (0.009) | **0.763 (0.001)**† | 0.855 (0.008) | 0.812 (0.004) | 0.885 (0.02) | 0.951 (0.02) |
| WFG6 m4 | 2.65 (0.003) | 2.44 (0.003)† | **2.43 (0.005)**† | 2.54 (0.008) | 2.52 (0.007) | 2.49 (0.003) | 2.47 (0.006) |
| ZDT1 d20 m2 | 1.40 (0.02) | 0.655 (0.08)† | **0.453 (0.04)**† | 0.654 (0.06) | 0.760 (0.02) | 0.608 (0.06)† | 0.730 (0.07) |
| ZDT2 d20 m2 | 1.63 (0.01) | 1.12 (0.008) | **0.700 (0.03)**† | 0.903 (0.07) | 1.15 (0.04) | 0.974 (0.06) | 1.20 (0.06) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.77 (0.01) | **1.61 (0.02)**† | 1.93 (0.03) | 2.00 (0.008) | 1.97 (0.02) | 1.89 (0.008) |
| DTLZ1 d15 m2 | 5.13 (0.01) | 4.60 (0.07)† | 4.43 (0.07)† | 4.50 (0.06)† | 4.65 (0.04) | **4.39 (0.03)**† | 4.45 (0.05)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.48 (0.1) | 6.42 (0.1) | 6.62 (0.08) | 6.76 (0.08) | **6.18 (0.02)**† | 6.23 (0.03)† |
| DTLZ1 d20 m2 | 5.20 (0.003) | 4.66 (0.04)† | 4.66 (0.03)† | 4.73 (0.04)† | 4.82 (0.06) | **4.64 (0.02)**† | 4.72 (0.03)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.76 (0.03)† | 6.71 (0.02)† | 6.99 (0.04) | 7.25 (0.06) | **6.68 (0.02)**† | 6.74 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.673 (0.03)† | **-0.735 (0.02)**† | -0.539 (0.03) | -0.372 (0) | -0.586 (0.005) | -0.372 (0) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.214 (0.03)† | **-0.272 (0.02)**† | -0.159 (0.01) | -0.0969 (0.003) | -0.190 (0.002) | -0.134 (0.02) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.562 (0.03)† | -0.553 (0.04)† | -0.426 (0.02) | -0.372 (0) | **-0.583 (0.004)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.129 (0.02) | -0.178 (0.02)† | -0.111 (0.009) | -0.0929 (0) | **-0.186 (0.002)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003)† | 0.108 (0.02)† | **0.0768 (0.03)**† | 0.150 (0.005)† | 0.146 (0.005)† | 0.140 (0.005)† | 0.0989 (0.008)† |
| Penicillin | 5.71 (0.01) | 5.64 (0.03)† | 5.65 (0.05)† | 5.67 (0.01)† | 5.67 (0.009)† | **5.63 (0.03)**† | 5.64 (0.04)† |
| VehicleSafety | 1.21 (0.01) | -0.0573 (0.04) | **-0.391 (0.04)**† | 0.375 (0.04) | 0.325 (0.03) | 0.473 (0.06) | 0.364 (0.04) |
| WFG6 m4 | 2.65 (0.003) | **2.14 (0.004)**† | **2.14 (0.01)**† | 2.42 (0.02) | 2.35 (0.008) | 2.21 (0.009) | 2.15 (0.01)† |
| ZDT1 d20 m2 | 1.41 (0.02) | 0.376 (0.05)† | 0.288 (0.06)† | 0.673 (0.06) | 0.579 (0.03) | **0.0335 (0.2)**† | 0.440 (0.1)† |
| ZDT2 d20 m2 | 1.63 (0.01) | -0.673 (0.2)† | **-0.931 (0.07)**† | 0.633 (0.1) | 1.22 (0.02) | 0.608 (0.1) | 1.00 (0.1) |

Table 22: Terminal $\log_{10}$ HV gap at $\eta = 0.2$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.31 (0.01) | 2.06 (0.01) | **1.95 (0.03)**† | 2.17 (0.03) | 1.99 (0.02)† | 1.96 (0.01)† | 2.04 (0.02)† |
| DTLZ1 d15 m2 | 5.16 (0.02) | 4.89 (0.04)† | 4.95 (0.04)† | 5.02 (0.03) | 4.92 (0.05)† | **4.78 (0.08)**† | 4.96 (0.04)† |
| DTLZ1 d15 m3 | 7.64 (0.01) | 7.12 (0.07)† | 7.13 (0.08) | 7.37 (0.04) | 7.20 (0.06) | **6.84 (0.09)**† | 6.96 (0.08)† |
| DTLZ1 d20 m2 | 5.20 (0) | 5.02 (0.03) | 5.02 (0.03)† | 5.09 (0.03) | **4.95 (0.02)**† | 4.99 (0.05)† | 5.06 (0.03) |
| DTLZ1 d20 m3 | 7.78 (0.005) | 7.44 (0.07) | 7.53 (0.07) | 7.62 (0.03) | 7.39 (0.05) | **7.19 (0.03)**† | 7.22 (0.05)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (0) | -0.374 (0.002) | -0.372 (0) | -0.376 (0.003) | **-0.542 (0.02)**† | -0.469 (0.03)† |
| DTLZ2 d15 m3 | -0.0938 (9e-4) | -0.0963 (0.003) | -0.0991 (0.005) | -0.0994 (0.006) | -0.0954 (0.001) | **-0.147 (0.01)**† | -0.142 (0.02)† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.376 (0.004) | **-0.533 (0.03)**† | -0.398 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0943 (8e-4)† | -0.0933 (4e-4)† | **-0.125 (0.01)**† | -0.112 (0.01)† |
| DTLZ2 m5 | 0.150 (0.003) | 0.146 (0.004) | 0.147 (0.005) | 0.123 (0.007)† | **0.117 (0.008)**† | 0.137 (0.006)† | 0.123 (0.008)† |
| Penicillin | 5.70 (0.007)† | 5.72 (0.01)† | 5.70 (0.009)† | 5.73 (0.02)† | 5.74 (0.02)† | **5.69 (0.02)**† | **5.69 (0.02)**† |
| VehicleSafety | 1.23 (0.02) | 0.562 (0.08) | **0.268 (0.05)**† | 0.606 (0.03) | 0.364 (0.04)† | 0.579 (0.06) | 0.577 (0.06) |
| WFG6 m4 | 2.68 (0.003) | 2.52 (0.04) | 2.56 (0.03) | 2.60 (0.005) | 2.47 (0.03) | **2.28 (0.01)**† | 2.32 (0.03)† |
| ZDT1 d20 m2 | 1.53 (0.01) | **1.06 (0.07)**† | 1.14 (0.1)† | 1.32 (0.04) | 1.34 (0.05) | 1.25 (0.06)† | 1.30 (0.06) |
| ZDT2 d20 m2 | 1.67 (0.01) | **1.05 (0.2)**† | 1.21 (0.07)† | 1.44 (0.04) | 1.25 (0.05)† | 1.46 (0.05) | 1.40 (0.06)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.29 (0.01) | 2.01 (0.03)† | **1.96 (0.02)**† | 2.08 (0.03) | 1.99 (0.02)† | **1.96 (0.01)**† | 2.02 (0.02) |
| DTLZ1 d15 m2 | 5.17 (0.02) | 4.92 (0.06)† | 4.99 (0.04) | 4.93 (0.06)† | 4.87 (0.05)† | 4.85 (0.03)† | **4.75 (0.07)**† |
| DTLZ1 d15 m3 | 7.63 (0.02) | 7.00 (0.1)† | 6.81 (0.07)† | 7.16 (0.09) | **6.76 (0.08)**† | 6.86 (0.08)† | 6.96 (0.1)† |
| DTLZ1 d20 m2 | 5.20 (0) | 5.01 (0.05)† | 5.04 (0.03)† | **4.98 (0.04)**† | 5.01 (0.04)† | **4.98 (0.06)**† | **4.98 (0.05)**† |
| DTLZ1 d20 m3 | 7.78 (0.006) | 7.45 (0.04) | 7.40 (0.1)† | 7.55 (0.06) | 7.35 (0.07)† | **7.21 (0.06)**† | **7.21 (0.06)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.378 (0.004) | -0.384 (0.009) | -0.385 (0.01) | -0.373 (6e-4) | **-0.538 (0.03)**† | -0.422 (0.03) |
| DTLZ2 d15 m3 | -0.0940 (9e-4) | -0.0947 (0.001) | -0.0963 (0.002) | -0.104 (0.009) | -0.0975 (0.002) | **-0.149 (0.01)**† | -0.122 (0.01)† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.375 (0.003) | -0.372 (0) | **-0.546 (0.02)**† | -0.415 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0933 (4e-4)† | -0.0929 (0)† | -0.0938 (6e-4)† | -0.0947 (0.002)† | **-0.108 (0.01)**† | -0.0929 (0)† |
| DTLZ2 m5 | 0.150 (0.003) | 0.146 (0.005) | 0.145 (0.004) | 0.125 (0.006)† | **0.119 (0.007)**† | 0.143 (0.005) | 0.124 (0.008)† |
| Penicillin | 5.70 (0.008)† | 5.73 (0.02) | 5.67 (0.01)† | 5.70 (0.01)† | 5.72 (0.02)† | 5.67 (0.01)† | **5.66 (0.02)**† |
| VehicleSafety | 1.22 (0.01) | 0.428 (0.06) | **0.127 (0.03)**† | 0.456 (0.06) | 0.375 (0.05) | 0.566 (0.06) | 0.487 (0.03) |
| WFG6 m4 | 2.68 (0.006) | 2.50 (0.03) | 2.44 (0.04) | 2.60 (0.01) | 2.45 (0.03) | **2.28 (0.02)**† | 2.32 (0.02)† |
| ZDT1 d20 m2 | 1.57 (0.02) | 1.00 (0.07)† | **0.660 (0.1)**† | 1.03 (0.07) | 0.931 (0.09)† | 1.15 (0.07) | 1.28 (0.04) |
| ZDT2 d20 m2 | 1.69 (0.009) | 0.702 (0.2)† | **0.360 (0.2)**† | 1.22 (0.06) | 1.14 (0.05) | 1.40 (0.06) | 1.44 (0.03) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.29 (0.01) | 1.99 (0.01) | **1.91 (0.02)**† | 1.95 (0.02)† | 2.10 (0.01) | 1.98 (0.007) | 2.01 (0.02) |
| DTLZ1 d15 m2 | 5.16 (0.02) | 5.02 (0.05) | 4.91 (0.05)† | 4.87 (0.06)† | 4.99 (0.07)† | 4.79 (0.05)† | **4.78 (0.07)**† |
| DTLZ1 d15 m3 | 7.69 (0.02) | 7.41 (0.07)† | 7.32 (0.09)† | 7.30 (0.06)† | 7.38 (0.09)† | **7.27 (0.09)**† | 7.31 (0.08)† |
| DTLZ1 d20 m2 | 5.20 (4e-4) | 5.07 (0.03)† | 5.03 (0.06)† | 5.04 (0.03)† | 5.17 (0.03) | **5.01 (0.04)**† | 5.05 (0.04)† |
| DTLZ1 d20 m3 | 7.80 (0.005) | 7.63 (0.04) | 7.68 (0.05) | 7.65 (0.07) | 7.61 (0.08) | 7.44 (0.05)† | **7.40 (0.05)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.376 (0.003) | -0.372 (0) | -0.377 (0.005) | -0.372 (7e-5) | **-0.561 (0.02)**† | -0.481 (0.03) |
| DTLZ2 d15 m3 | -0.0957 (0.002) | -0.0976 (0.002) | -0.0960 (0.002) | -0.105 (0.007) | -0.104 (0.005) | -0.187 (0.003)† | **-0.193 (0.007)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.376 (0.004) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.527 (0.03)**† | -0.406 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0945 (0.001) | -0.0929 (0) | -0.0949 (0.002) | **-0.177 (0.003)**† | -0.168 (0.01)† |
| DTLZ2 m5 | 0.144 (0.002)† | 0.147 (0.003) | 0.145 (0.005)† | 0.128 (0.007)† | **0.126 (0.009)**† | 0.151 (0.003) | 0.135 (0.008)† |
| Penicillin | 5.71 (0.02) | **5.61 (0.03)**† | 5.66 (0.03)† | 5.69 (0.03)† | 5.62 (0.02)† | 5.63 (0.02)† | 5.64 (0.02)† |
| VehicleSafety | 1.26 (0.01) | 0.955 (0.03) | **0.795 (0.004)**† | 0.853 (0.02) | 0.829 (0.007) | 0.920 (0.02) | 0.932 (0.02) |
| WFG6 m4 | 2.72 (0.01) | 2.59 (0.02)† | 2.62 (0.01) | 2.61 (0.01) | 2.61 (0.02)† | 2.61 (0.02)† | **2.58 (0.01)**† |
| ZDT1 d20 m2 | 1.46 (0.02) | 1.12 (0.05)† | **1.11 (0.06)**† | 1.23 (0.07)† | 1.14 (0.07)† | 1.15 (0.06)† | 1.22 (0.05)† |
| ZDT2 d20 m2 | 1.65 (0.01) | 1.30 (0.04)† | **1.24 (0.1)**† | 1.29 (0.03)† | **1.24 (0.03)**† | 1.33 (0.04)† | 1.40 (0.03) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.30 (0.009) | 2.11 (0.01) | 2.07 (0.02)† | 2.11 (0.04) | 2.09 (0.009) | **2.00 (0.02)**† | 2.03 (0.02)† |
| DTLZ1 d15 m2 | 5.16 (0.02) | 4.88 (0.05) | 4.89 (0.06) | 4.89 (0.07) | 5.13 (0.02) | **4.52 (0.04)**† | 4.61 (0.05)† |
| DTLZ1 d15 m3 | 7.58 (0.03) | 7.22 (0.07) | 7.15 (0.08) | 7.35 (0.06) | 7.59 (0.02) | **6.38 (0.05)**† | 6.45 (0.04)† |
| DTLZ1 d20 m2 | 5.20 (5e-4) | 5.10 (0.04) | 5.08 (0.04) | 5.02 (0.04) | 5.20 (0.006) | **4.76 (0.03)**† | 4.78 (0.05)† |
| DTLZ1 d20 m3 | 7.77 (0.01) | 7.51 (0.06) | 7.57 (0.05) | 7.65 (0.05) | 7.77 (0.01) | **6.84 (0.04)**† | 6.99 (0.07)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (2e-5) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.501 (0.03)**† | -0.454 (0.03)† |
| DTLZ2 d15 m3 | -0.0929 (0) | -0.0932 (3e-4) | -0.0932 (3e-4) | -0.0966 (0.004) | -0.0938 (8e-4) | -0.121 (0.01)† | **-0.165 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.487 (0.03)**† | -0.389 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0934 (5e-4)† | -0.0931 (1e-4)† | -0.101 (0.008)† | **-0.128 (0.01)**† |
| DTLZ2 m5 | 0.150 (0.002)† | 0.150 (0.004)† | 0.153 (0.004)† | **0.139 (0.009)**† | **0.139 (0.006)**† | 0.158 (0.002)† | 0.149 (0.006)† |
| Penicillin | 5.71 (0.01) | 5.70 (0.05)† | 5.77 (0.05)† | 5.70 (0.02) | 5.71 (0.02) | **5.64 (0.02)**† | 5.74 (0.05)† |
| VehicleSafety | 1.24 (0.02) | 0.559 (0.06) | **0.336 (0.06)**† | 0.586 (0.06) | 0.396 (0.05)† | 0.510 (0.03) | 0.530 (0.05) |
| WFG6 m4 | 2.67 (0.003) | 2.63 (0.01) | 2.62 (0.01) | 2.62 (0.01) | 2.64 (0.01) | 2.31 (0.02)† | **2.29 (0.02)**† |
| ZDT1 d20 m2 | 1.51 (0.04) | 1.10 (0.1)† | 1.05 (0.1)† | 1.30 (0.1) | 1.51 (0.01) | **0.777 (0.1)**† | 0.998 (0.1)† |
| ZDT2 d20 m2 | 1.66 (0.02) | 1.02 (0.1)† | **0.804 (0.2)**† | 1.10 (0.07) | 1.48 (0.04) | 1.21 (0.07) | 1.31 (0.05) |

Table 23: Terminal $\log_{10}$ HV gap at $\eta = 0.4$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.35 (0.01) | 2.25 (0.03) | 2.22 (0.03) | 2.26 (0.03) | 2.25 (0.04) | **2.10 (0.03)**† | 2.19 (0.02) |
| DTLZ1 d15 m2 | 5.18 (0.02) | 4.99 (0.05)† | 5.10 (0.04) | 5.14 (0.03) | 5.13 (0.04) | **4.88 (0.05)**† | 4.89 (0.05)† |
| DTLZ1 d15 m3 | 7.67 (0.02) | 7.50 (0.06) | 7.53 (0.04) | 7.57 (0.05) | 7.44 (0.06) | **6.96 (0.1)**† | 7.15 (0.09)† |
| DTLZ1 d20 m2 | 5.20 (0) | 5.13 (0.03) | 5.20 (0.006) | 5.20 (0) | 5.17 (0.02) | **4.95 (0.04)**† | 4.97 (0.02)† |
| DTLZ1 d20 m3 | 7.79 (0.005) | 7.74 (0.02) | 7.77 (0.01) | 7.74 (0.03) | 7.74 (0.04) | **7.34 (0.07)**† | 7.39 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0)† | -0.372 (1e-5)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | **-0.386 (0.01)**† | -0.384 (0.01)† |
| DTLZ2 d15 m3 | -0.0938 (9e-4) | -0.0947 (0.001)† | -0.0931 (2e-4) | -0.0934 (3e-4)† | -0.0993 (0.003)† | -0.114 (0.01)† | **-0.126 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | **-0.391 (0.02)**† | -0.386 (0.01)† |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0929 (0)† | -0.0933 (4e-4)† | -0.0929 (0)† | -0.0929 (0)† | -0.0975 (0.003)† | **-0.100 (0.007)**† |
| DTLZ2 m5 | 0.154 (0.002) | 0.130 (0.008)† | 0.141 (0.005) | **0.117 (0.006)**† | 0.126 (0.008)† | 0.144 (0.006) | 0.138 (0.008)† |
| Penicillin | **5.74 (0.01)**† | 5.77 (0.02)† | 5.77 (0.02)† | 5.76 (0.02)† | 5.80 (0.05)† | 5.78 (0.02)† | 5.77 (0.02)† |
| VehicleSafety | 1.31 (0.02) | 0.935 (0.09) | **0.671 (0.08)**† | 0.940 (0.09) | 0.847 (0.08)† | 0.890 (0.05) | 0.808 (0.07)† |
| WFG6 m4 | 2.70 (0.003) | 2.66 (0.01) | 2.68 (0.007) | 2.70 (0.007) | 2.65 (0.03) | **2.36 (0.01)**† | 2.40 (0.03)† |
| ZDT1 d20 m2 | 1.57 (0.02)† | 1.58 (0.03) | 1.58 (0.03) | 1.65 (0.03) | 1.66 (0.02) | **1.46 (0.04)**† | 1.47 (0.05)† |
| ZDT2 d20 m2 | 1.71 (0.02) | 1.61 (0.03)† | **1.55 (0.03)**† | 1.63 (0.02) | 1.60 (0.04)† | 1.57 (0.04)† | 1.56 (0.04)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.34 (0.01) | 2.18 (0.05) | 2.22 (0.03) | 2.21 (0.04) | 2.13 (0.04) | **2.00 (0.03)**† | 2.13 (0.02) |
| DTLZ1 d15 m2 | 5.18 (0.02) | 4.99 (0.06)† | 5.06 (0.04) | 5.03 (0.05)† | 4.91 (0.09)† | 4.87 (0.03)† | **4.86 (0.05)**† |
| DTLZ1 d15 m3 | 7.67 (0.02) | 7.33 (0.07) | 7.34 (0.05) | 7.43 (0.04) | **6.91 (0.06)**† | 6.98 (0.08)† | 7.22 (0.1) |
| DTLZ1 d20 m2 | 5.20 (0.006) | 5.12 (0.03) | 5.14 (0.03) | 5.19 (0.02) | 5.09 (0.03)† | 5.02 (0.03)† | **5.00 (0.04)**† |
| DTLZ1 d20 m3 | 7.79 (0.005) | 7.63 (0.06) | 7.70 (0.05) | 7.74 (0.02) | 7.40 (0.08)† | 7.33 (0.05)† | **7.29 (0.06)**† |
| DTLZ2 d15 m2 | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | **-0.410 (0.03)**† |
| DTLZ2 d15 m3 | -0.0931 (2e-4)† | -0.0934 (5e-4)† | -0.0961 (0.001)† | -0.0933 (4e-4)† | -0.0945 (0.001)† | -0.111 (0.01)† | **-0.123 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | **-0.402 (0.02)**† | -0.390 (0.02)† |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0929 (0)† | -0.0929 (0)† | -0.0932 (2e-4)† | -0.0934 (5e-4)† | **-0.107 (0.01)**† | -0.102 (0.009)† |
| DTLZ2 m5 | 0.151 (0.003)† | 0.142 (0.005)† | 0.144 (0.006)† | **0.128 (0.009)**† | 0.134 (0.005)† | 0.152 (0.004) | 0.137 (0.007)† |
| Penicillin | 5.80 (0.05) | 5.72 (0.01)† | 5.71 (0.02)† | **5.68 (0.02)**† | 5.75 (0.02)† | 5.72 (0.02)† | 5.71 (0.02)† |
| VehicleSafety | 1.25 (0.01) | 0.650 (0.07)† | **0.590 (0.09)**† | 0.826 (0.08)† | 0.663 (0.08)† | 0.809 (0.05) | 0.657 (0.06)† |
| WFG6 m4 | 2.70 (0.007) | 2.68 (0.01) | 2.64 (0.03) | 2.66 (0.01) | 2.49 (0.04) | 2.35 (0.02)† | **2.34 (0.02)**† |
| ZDT1 d20 m2 | 1.56 (0.01) | 1.42 (0.08) | **1.25 (0.04)**† | 1.41 (0.07)† | 1.28 (0.05)† | 1.52 (0.04) | 1.48 (0.05) |
| ZDT2 d20 m2 | 1.69 (0.008) | 1.46 (0.04) | **1.27 (0.03)**† | 1.44 (0.04) | 1.39 (0.04)† | 1.58 (0.02) | 1.62 (0.03) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.31 (0.008) | 2.10 (0.02)† | **2.05 (0.02)**† | 2.10 (0.03)† | 2.19 (0.03) | 2.06 (0.02)† | 2.11 (0.02) |
| DTLZ1 d15 m2 | 5.18 (0.01) | 4.94 (0.05)† | 5.05 (0.04)† | 5.09 (0.04) | 5.16 (0.02) | **4.89 (0.06)**† | 5.04 (0.03)† |
| DTLZ1 d15 m3 | 7.66 (0.03) | 7.50 (0.08) | 7.63 (0.06) | 7.64 (0.06) | 7.73 (0.04) | **6.95 (0.08)**† | 7.19 (0.09)† |
| DTLZ1 d20 m2 | 5.20 (5e-4) | 5.16 (0.02) | 5.14 (0.03)† | 5.18 (0.02) | 5.20 (0.001) | **5.01 (0.06)**† | 5.03 (0.04)† |
| DTLZ1 d20 m3 | 7.79 (0.006) | 7.65 (0.06) | 7.73 (0.03) | 7.78 (0.01) | 7.73 (0.06) | **7.25 (0.05)**† | 7.30 (0.07)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.375 (0.003) | -0.372 (0) | -0.373 (8e-4) | -0.372 (0) | **-0.543 (0.02)**† | -0.484 (0.03)† |
| DTLZ2 d15 m3 | -0.0933 (3e-4) | -0.102 (0.004)† | -0.0930 (1e-4)† | -0.102 (0.005)† | -0.0954 (0.001)† | -0.108 (0.01)† | **-0.130 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.455 (0.03)**† | -0.408 (0.02)† |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0957 (0.003)† | -0.0929 (1e-5)† | -0.0933 (4e-4)† | -0.0932 (3e-4)† | -0.101 (0.008)† | **-0.110 (0.01)**† |
| DTLZ2 m5 | 0.152 (0.002) | 0.131 (0.007)† | **0.129 (0.006)**† | 0.142 (0.003)† | **0.129 (0.005)**† | 0.154 (0.004) | 0.154 (0.003) |
| Penicillin | 5.73 (0.03) | 5.73 (0.08)† | 5.69 (0.02) | 5.63 (0.01)† | 5.65 (0.01)† | **5.61 (0.02)**† | 5.63 (0.02)† |
| VehicleSafety | 1.31 (0.02) | 0.968 (0.03) | **0.845 (0.009)**† | 0.868 (0.01)† | 0.897 (0.03) | 0.960 (0.03) | 0.972 (0.02) |
| WFG6 m4 | 2.78 (0.007) | 2.76 (0.007) | 2.75 (0.01) | 2.73 (0.02) | 2.72 (0.02) | 2.71 (0.02)† | **2.67 (0.02)**† |
| ZDT1 d20 m2 | 1.57 (0.03) | 1.55 (0.03) | **1.37 (0.06)**† | 1.56 (0.03) | 1.62 (0.02) | 1.43 (0.05)† | 1.51 (0.03)† |
| ZDT2 d20 m2 | 1.67 (0.01) | **1.50 (0.06)**† | 1.58 (0.04)† | 1.53 (0.03)† | 1.60 (0.02)† | 1.53 (0.02)† | 1.57 (0.03)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.35 (0.01) | 2.32 (0.02) | 2.28 (0.02) | 2.24 (0.02)† | 2.31 (0.02) | **2.15 (0.04)**† | 2.20 (0.03)† |
| DTLZ1 d15 m2 | 5.18 (0.007) | 5.09 (0.04) | 5.09 (0.04) | 5.14 (0.03) | 5.19 (0.006) | **4.76 (0.09)**† | 4.82 (0.06)† |
| DTLZ1 d15 m3 | 7.63 (0.03) | 7.56 (0.04) | 7.51 (0.03) | 7.60 (0.04) | 7.71 (0.03) | **6.72 (0.1)**† | 6.93 (0.1)† |
| DTLZ1 d20 m2 | 5.20 (5e-4) | 5.19 (0.01) | 5.20 (0.004) | 5.19 (0.01) | 5.20 (8e-4) | **4.96 (0.05)**† | **4.96 (0.05)**† |
| DTLZ1 d20 m3 | 7.79 (0.01) | 7.77 (0.01) | 7.75 (0.02) | 7.79 (0.006) | 7.79 (0.01) | **7.20 (0.07)**† | **7.20 (0.06)**† |
| DTLZ2 d15 m2 | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | **-0.382 (0.01)**† |
| DTLZ2 d15 m3 | -0.0929 (0)† | -0.0929 (0)† | -0.0929 (0)† | -0.0939 (8e-4)† | -0.0931 (2e-4)† | -0.0983 (0.005)† | **-0.104 (0.008)**† |
| DTLZ2 d20 m2 | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | -0.372 (0)† | **-0.387 (0.01)**† |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0929 (0)† | -0.0929 (0)† | **-0.0936 (5e-4)**† | -0.0929 (0)† | -0.0929 (0)† | -0.0929 (0)† |
| DTLZ2 m5 | 0.155 (0.002) | 0.155 (0.002) | 0.153 (0.003)† | **0.125 (0.01)**† | 0.146 (0.004)† | 0.158 (0.001) | 0.160 (4e-4) |
| Penicillin | 5.74 (0.01) | 5.74 (0.04)† | 5.80 (0.03)† | 5.77 (0.02) | 5.75 (0.02) | **5.68 (0.02)**† | 5.74 (0.03)† |
| VehicleSafety | 1.30 (0.02) | 1.02 (0.09) | 0.841 (0.09)† | 0.758 (0.06)† | 1.13 (0.05) | **0.756 (0.07)**† | 0.926 (0.09)† |
| WFG6 m4 | 2.68 (0.003) | 2.70 (0.01) | 2.70 (0.01) | 2.68 (0.01) | 2.68 (0.01) | **2.37 (0.02)**† | 2.38 (0.02)† |
| ZDT1 d20 m2 | 1.52 (0.04) | 1.44 (0.08)† | 1.52 (0.06) | 1.63 (0.03) | 1.52 (0.04) | **1.30 (0.05)**† | 1.49 (0.07)† |
| ZDT2 d20 m2 | 1.67 (0.02) | 1.64 (0.04) | 1.59 (0.05)† | 1.64 (0.02) | 1.69 (0.008) | **1.31 (0.1)**† | 1.57 (0.04)† |

Table 24: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.05$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0192 (8e-4)† | **0.0167 (0.001)**† | 0.0497 (0.006) | 0.0632 (0.001) | 0.0375 (0.002) | 0.0597 (0.003) |
| DTLZ1 d15 m2 | 756 (30) | 290 (10)† | 323 (30)† | 291 (20)† | 305 (10)† | **276 (10)**† | 315 (20)† |
| DTLZ1 d15 m3 | 512 (30) | 229 (6)† | 230 (8) | **209 (6)**† | 263 (10) | 223 (7)† | 227 (3) |
| DTLZ1 d20 m2 | 1104 (40) | **378 (10)**† | 396 (30)† | 393 (20)† | 438 (20)† | 429 (20) | 472 (30) |
| DTLZ1 d20 m3 | 775 (20) | 304 (8)† | **297 (6)**† | 310 (20)† | 367 (20) | 303 (9)† | 313 (7)† |
| DTLZ2 d15 m2 | 0.741 (0.02) | **0.193 (0.01)**† | 0.222 (0.02)† | 0.195 (0.02)† | 0.824 (0.04) | 0.317 (0.005) | 0.664 (0.07) |
| DTLZ2 d15 m3 | 0.702 (0.02) | 0.526 (0.07)† | 0.471 (0.06)† | 0.554 (0.06)† | 0.762 (0.02) | **0.410 (0.01)**† | 0.762 (0.04) |
| DTLZ2 d20 m2 | 1.05 (0.02) | 0.343 (0.03)† | 0.438 (0.1)† | 0.365 (0.03)† | 1.19 (0.05) | **0.329 (0.002)**† | 1.12 (0.03) |
| DTLZ2 d20 m3 | 0.994 (0.02)† | **0.656 (0.1)**† | 0.799 (0.1)† | 0.739 (0.05)† | 1.06 (0.03) | 0.764 (0.1)† | 1.08 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.482 (0.05)† | **0.463 (0.04)**† | 0.564 (0.03) | 0.707 (0.02) | 0.526 (0.01) | 0.473 (0.003) |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 0.00227 (7e-4) | 1.28e-4 (1e-4)† | 3.24e-5 (2e-5) |
| Penicillin | 0.0827 (0.007)† | **0.0743 (0.006)**† | 0.0757 (0.004)† | 0.106 (0.01) | 0.0785 (0.004)† | 0.0874 (0.005)† | 0.0899 (0.003) |
| VehicleSafety | 0.181 (0.009) | 0.00486 (1e-3) | **6.80e-4 (1e-4)**† | 0.0126 (0.001) | 0.00963 (0.001) | 0.0119 (0.002) | 0.0135 (0.002) |
| WFG6 m4 | 0.124 (0.002) | **0.0372 (0.002)**† | 0.0419 (0.003)† | 0.0881 (0.004) | 0.0735 (0.005) | 0.0569 (0.002) | 0.0465 (0.003) |
| ZDT1 d20 m2 | 2.02 (0.09) | **0.0911 (0.02)**† | 0.129 (0.02)† | 0.131 (0.03)† | 0.360 (0.02) | 0.452 (0.06) | 0.395 (0.08) |
| ZDT2 d20 m2 | 3.23 (0.08) | **0.0711 (0.02)**† | 0.0849 (0.01)† | 0.111 (0.03)† | 0.747 (0.06) | 0.751 (0.2) | 0.604 (0.1) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0193 (8e-4)† | **0.0183 (0.002)**† | 0.0481 (0.005) | 0.0668 (0.001) | 0.0424 (0.004) | 0.0552 (0.003) |
| DTLZ1 d15 m2 | 772 (40) | 285 (10)† | 283 (10)† | 365 (90)† | 340 (20) | **273 (10)**† | 292 (10)† |
| DTLZ1 d15 m3 | 514 (30) | 222 (7)† | 236 (7) | **206 (6)**† | 249 (10) | 236 (10) | 227 (6) |
| DTLZ1 d20 m2 | 1045 (50) | 373 (20)† | 386 (20)† | 400 (30)† | 404 (20)† | **372 (20)**† | 400 (10)† |
| DTLZ1 d20 m3 | 775 (20) | 293 (7)† | **288 (8)**† | 300 (10)† | 384 (10) | 326 (8) | 325 (10) |
| DTLZ2 d15 m2 | 0.726 (0.02) | 0.266 (0.03)† | **0.242 (0.03)**† | 0.288 (0.04)† | 0.829 (0.02) | 0.323 (0.005) | 0.778 (0.02) |
| DTLZ2 d15 m3 | 0.711 (0.02) | 0.481 (0.07)† | 0.469 (0.06)† | 0.448 (0.04)† | 0.767 (0.02) | **0.405 (0.005)**† | 0.792 (0.02) |
| DTLZ2 d20 m2 | 1.03 (0.02) | 0.348 (0.03)† | **0.320 (0.04)**† | 0.350 (0.04)† | 1.12 (0.03) | 0.332 (0.005)† | 1.12 (0.03) |
| DTLZ2 d20 m3 | 0.992 (0.02)† | 0.795 (0.1)† | 0.798 (0.1)† | 0.707 (0.06)† | 1.06 (0.02) | **0.640 (0.1)**† | 1.08 (0.02) |
| DTLZ2 m5 | 0.645 (0.01) | 0.484 (0.05)† | 0.484 (0.05)† | 0.545 (0.03)† | 0.726 (0.03) | 0.568 (0.04) | **0.465 (0.006)**† |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 9.92e-4 (4e-4) | 3.02e-5 (2e-5) | 5.69e-5 (3e-5) |
| Penicillin | 0.0835 (0.005)† | 0.0869 (0.008)† | **0.0768 (0.004)**† | 0.147 (0.03) | 0.0919 (0.008)† | 0.0914 (0.005) | 0.0929 (0.005)† |
| VehicleSafety | 0.182 (0.009) | 0.00337 (0.001) | **8.20e-4 (1e-4)**† | 0.0118 (8e-4) | 0.00874 (7e-4) | 0.0200 (0.005) | 0.0221 (0.005) |
| WFG6 m4 | 0.124 (0.002) | **0.0342 (0.002)**† | 0.0354 (0.002)† | 0.0918 (0.004) | 0.0724 (0.006) | 0.0560 (0.002) | 0.0439 (0.002) |
| ZDT1 d20 m2 | 2.03 (0.09) | **0.0835 (0.01)**† | 0.0908 (0.01)† | 0.101 (0.02)† | 0.385 (0.02) | 0.416 (0.1) | 0.520 (0.08) |
| ZDT2 d20 m2 | 3.24 (0.09) | **0.0457 (0.005)**† | 0.0865 (0.01) | 0.231 (0.03) | 0.805 (0.06) | 0.587 (0.09) | 0.830 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.107 (0.002) | 0.0357 (0.002) | **0.0287 (0.001)**† | 0.0575 (0.004) | 0.0769 (0.001) | 0.0511 (0.002) | 0.0588 (0.002) |
| DTLZ1 d15 m2 | 720 (50) | 342 (10)† | 339 (20)† | 321 (10)† | 414 (50) | **316 (9)**† | 324 (10)† |
| DTLZ1 d15 m3 | 523 (30) | 282 (20)† | 264 (7) | **243 (4)**† | 294 (10) | 266 (9)† | 279 (7) |
| DTLZ1 d20 m2 | 1063 (20) | 451 (20)† | 466 (30)† | 531 (103)† | 624 (123)† | **443 (20)**† | 513 (30)† |
| DTLZ1 d20 m3 | 831 (30) | 365 (20)† | 333 (7)† | **328 (9)**† | 434 (30) | 342 (8)† | 402 (10) |
| DTLZ2 d15 m2 | 0.767 (0.03) | 0.251 (0.03)† | **0.222 (0.01)**† | 0.231 (0.02)† | 0.749 (0.03) | 0.360 (0.03) | 0.686 (0.05) |
| DTLZ2 d15 m3 | 0.641 (0.02) | 0.433 (0.05)† | 0.437 (0.05)† | 0.452 (0.03)† | 0.669 (0.03) | **0.394 (0.007)**† | 0.576 (0.06) |
| DTLZ2 d20 m2 | 1.04 (0.02) | 0.704 (0.09) | 0.519 (0.1)† | 0.414 (0.05)† | 1.11 (0.02) | **0.341 (0.005)**† | 1.07 (0.03) |
| DTLZ2 d20 m3 | 0.974 (0.02) | 0.715 (0.1)† | 0.852 (0.08) | 0.839 (0.06) | 0.969 (0.04) | **0.413 (0.003)**† | 1.06 (0.01) |
| DTLZ2 m5 | 0.608 (0.008) | 0.542 (0.04)† | 0.639 (0.07)† | 0.618 (0.04) | 0.715 (0.05) | 1.02 (0.2) | **0.447 (0.007)**† |
| DTLZ7 m5 | 0.206 (0.008) | 5.47e-6 (2e-6) | **0 (0)**† | 5.15e-8 (5e-8)† | 0.0180 (0.005) | 3.98e-8 (4e-8)† | 5.81e-5 (5e-5) |
| Penicillin | 0.0897 (0.007)† | 0.150 (0.05)† | **0.0753 (0.003)**† | 0.129 (0.03)† | 0.0966 (0.008)† | 0.110 (0.007) | 0.0870 (0.004) |
| VehicleSafety | 0.189 (0.01) | 0.0268 (8e-4) | **0.0169 (7e-5)**† | 0.0289 (1e-3) | 0.0237 (9e-4) | 0.0357 (0.003) | 0.0457 (0.004) |
| WFG6 m4 | 0.125 (0.002) | 0.0847 (0.002)† | **0.0833 (0.003)**† | 0.111 (0.002) | 0.113 (0.003) | 0.109 (0.002) | 0.0988 (0.002) |
| ZDT1 d20 m2 | 1.98 (0.08) | 0.421 (0.04) | **0.194 (0.02)**† | 0.250 (0.04)† | 0.448 (0.02) | 0.361 (0.04) | 0.425 (0.04) |
| ZDT2 d20 m2 | 3.25 (0.09) | 0.595 (0.02) | **0.279 (0.008)**† | 0.380 (0.05)† | 0.861 (0.07) | 0.492 (0.07) | 0.938 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0297 (0.002) | **0.0197 (0.002)**† | 0.0569 (0.004) | 0.0616 (0.002) | 0.0313 (0.001) | 0.0513 (0.002) |
| DTLZ1 d15 m2 | 737 (30) | 278 (10) | 260 (10)† | 262 (10)† | 298 (10) | 249 (1)† | **244 (2)**† |
| DTLZ1 d15 m3 | 512 (30) | 229 (10)† | 221 (10)† | 219 (10)† | 260 (10) | **201 (2)**† | 214 (6)† |
| DTLZ1 d20 m2 | 1023 (40) | 344 (5)† | 346 (7)† | 363 (20)† | 441 (80)† | 346 (9)† | **342 (5)**† |
| DTLZ1 d20 m3 | 775 (20) | 299 (9)† | **290 (5)**† | 298 (10)† | 380 (30) | 300 (9)† | 310 (10) |
| DTLZ2 d15 m2 | 0.769 (0.02) | 0.190 (0.03)† | **0.161 (0.009)**† | 0.247 (0.05) | 0.774 (0.03) | 0.319 (0.004) | 0.761 (0.02) |
| DTLZ2 d15 m3 | 0.711 (0.02) | 0.463 (0.08)† | **0.319 (0.03)**† | 0.434 (0.05) | 0.697 (0.03) | 0.393 (0.003) | 0.641 (0.07) |
| DTLZ2 d20 m2 | 1.11 (0.04) | 0.428 (0.1)† | 0.598 (0.2)† | 0.391 (0.06)† | 1.04 (0.05) | **0.322 (0.003)**† | 1.08 (0.02) |
| DTLZ2 d20 m3 | 0.996 (0.02) | 0.753 (0.1) | 0.589 (0.1)† | 0.705 (0.08) | 1.01 (0.03) | **0.401 (0.003)**† | 1.07 (0.02) |
| DTLZ2 m5 | 0.645 (0.01) | 0.750 (0.1)† | 0.680 (0.1)† | 0.646 (0.04) | 0.653 (0.02) | 0.959 (0.2) | **0.454 (0.009)**† |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 0.00285 (6e-4) | 2.34e-5 (2e-5)† | 3.79e-6 (4e-6)† |
| Penicillin | 0.0888 (0.008)† | 0.0899 (0.01)† | 0.116 (0.04)† | **0.0804 (0.007)**† | 0.0811 (0.007)† | 0.0968 (0.006) | 0.105 (0.009) |
| VehicleSafety | 0.182 (0.01) | 0.00339 (5e-4) | **7.23e-4 (1e-4)**† | 0.0118 (0.002) | 0.00996 (0.001) | 0.0146 (0.001) | 0.0110 (0.002) |
| WFG6 m4 | 0.124 (0.002) | 0.0413 (0.001)† | **0.0389 (0.002)**† | 0.0899 (0.004) | 0.0810 (0.003) | 0.0568 (0.002) | 0.0420 (0.001)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 0.146 (0.01) | **0.0875 (0.009)**† | 0.218 (0.03) | 0.333 (0.01) | 0.229 (0.05) | 0.337 (0.05) |
| ZDT2 d20 m2 | 3.28 (0.1) | 0.108 (0.04)† | **0.0578 (0.008)**† | 0.175 (0.04)† | 0.979 (0.07) | 0.349 (0.04) | 0.741 (0.1) |

Table 25: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.2$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.123 (0.006) | 0.0819 (0.005) | **0.0623 (0.006)**† | 0.100 (0.009) | 0.0846 (0.005) | 0.0697 (0.007)† | 0.0876 (0.008) |
| DTLZ1 d15 m2 | 813 (50) | **412 (70)**† | 458 (50)† | 527 (40) | 486 (70)† | 448 (100)† | 469 (50)† |
| DTLZ1 d15 m3 | 592 (20) | 300 (20)† | 294 (20)† | 412 (30) | 288 (10)† | 295 (10)† | **282 (20)**† |
| DTLZ1 d20 m2 | 1212 (60) | 535 (40)† | 536 (40)† | 656 (70) | **484 (10)**† | 559 (40)† | 618 (60) |
| DTLZ1 d20 m3 | 873 (30) | 458 (40)† | 530 (40) | 622 (40) | 429 (30)† | **383 (20)**† | 390 (10)† |
| DTLZ2 d15 m2 | 0.905 (0.04) | 1.11 (0.1) | 1.44 (0.3) | 1.23 (0.1) | 0.730 (0.05) | **0.445 (0.1)**† | 0.573 (0.09)† |
| DTLZ2 d15 m3 | 0.788 (0.03)† | 0.799 (0.04) | 0.738 (0.04)† | 0.675 (0.04)† | **0.564 (0.03)**† | **0.564 (0.07)**† | 0.613 (0.07)† |
| DTLZ2 d20 m2 | 1.14 (0.02) | 1.48 (0.2) | 1.75 (0.2) | 1.34 (0.2) | 1.10 (0.2) | **0.517 (0.1)**† | 1.06 (0.1) |
| DTLZ2 d20 m3 | 1.09 (0.02) | 1.08 (0.03) | 1.10 (0.04) | 0.881 (0.06) | **0.730 (0.03)**† | 0.853 (0.1)† | 0.900 (0.1)† |
| DTLZ2 m5 | 0.664 (0.01) | 0.650 (0.03) | 0.653 (0.02) | 0.553 (0.02)† | **0.516 (0.02)**† | 0.618 (0.04)† | 0.558 (0.04)† |
| DTLZ7 m5 | 0.200 (0.009) | 0.00684 (0.004)† | 0.00203 (0.001)† | 0.0271 (0.009) | 0.0174 (0.004) | **7.92e-4 (5e-4)**† | 0.00328 (0.001)† |
| Penicillin | **0.0841 (0.002)**† | 0.109 (0.008) | 0.100 (0.007)† | 0.120 (0.01) | 0.118 (0.01) | 0.106 (0.005) | 0.103 (0.005) |
| VehicleSafety | 0.189 (0.01) | 0.0302 (0.01) | **0.00904 (0.002)**† | 0.0226 (0.003) | 0.0147 (0.002) | 0.0247 (0.004) | 0.0252 (0.005) |
| WFG6 m4 | 0.145 (0.002) | 0.122 (0.008) | 0.127 (0.005) | 0.128 (0.003) | 0.111 (0.007) | 0.0816 (0.003)† | **0.0754 (0.003)**† |
| ZDT1 d20 m2 | 2.77 (0.1) | **0.921 (0.2)**† | 1.20 (0.3)† | 1.61 (0.2) | 1.63 (0.2) | 1.46 (0.2) | 1.68 (0.2) |
| ZDT2 d20 m2 | 3.62 (0.1) | **1.03 (0.2)**† | 1.08 (0.3)† | 1.99 (0.2) | 1.19 (0.2)† | 2.18 (0.3) | 1.88 (0.4)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.110 (0.004) | 0.0742 (0.005)† | **0.0635 (0.005)**† | 0.0767 (0.005)† | 0.0751 (0.006)† | 0.0658 (0.007)† | 0.0845 (0.007) |
| DTLZ1 d15 m2 | 855 (70) | 456 (60)† | 506 (40)† | 461 (50)† | 422 (50)† | 400 (20)† | **394 (30)**† |
| DTLZ1 d15 m3 | 580 (40) | 297 (30)† | **246 (10)**† | 346 (40) | 259 (10)† | 289 (10) | 274 (20)† |
| DTLZ1 d20 m2 | 1194 (40) | 606 (107)† | 635 (102)† | **516 (50)**† | 600 (80)† | 576 (40)† | 681 (184)† |
| DTLZ1 d20 m3 | 904 (40) | 457 (30)† | 452 (50)† | 559 (40) | 451 (30)† | 413 (20)† | **409 (10)**† |
| DTLZ2 d15 m2 | 1.05 (0.04) | 1.08 (0.2) | 0.954 (0.1) | 0.776 (0.08)† | 0.831 (0.08) | **0.536 (0.1)**† | 0.722 (0.08)† |
| DTLZ2 d15 m3 | 0.775 (0.03) | 0.751 (0.03) | 0.717 (0.03) | 0.713 (0.06) | **0.542 (0.03)**† | 0.562 (0.07)† | 0.662 (0.06)† |
| DTLZ2 d20 m2 | 1.27 (0.03) | 1.77 (0.2) | 1.36 (0.1) | 1.09 (0.2) | 0.946 (0.06) | **0.447 (0.1)**† | 1.11 (0.1) |
| DTLZ2 d20 m3 | 1.11 (0.02) | 1.04 (0.04) | 1.12 (0.02) | 0.853 (0.05)† | **0.750 (0.04)**† | 0.996 (0.1) | 1.07 (0.02) |
| DTLZ2 m5 | 0.681 (0.02) | 0.666 (0.02) | 0.641 (0.02) | 0.550 (0.02)† | **0.525 (0.03)**† | 0.685 (0.05) | 0.570 (0.04)† |
| DTLZ7 m5 | 0.200 (0.009) | 0.00263 (9e-4)† | **1.76e-4 (7e-5)**† | 0.0107 (0.006) | 0.0124 (0.003) | 4.21e-4 (2e-4)† | 0.00371 (0.002)† |
| Penicillin | **0.0850 (0.003)**† | 0.122 (0.02) | 0.0969 (0.006)† | 0.0913 (0.009)† | 0.109 (0.01) | 0.117 (0.01) | 0.109 (0.005) |
| VehicleSafety | 0.187 (0.009) | 0.0235 (0.008) | **0.00503 (8e-4)**† | 0.0152 (0.002) | 0.0164 (0.003) | 0.0216 (0.003) | 0.0188 (0.002) |
| WFG6 m4 | 0.145 (0.005) | 0.116 (0.007) | 0.102 (0.009) | 0.127 (0.006) | 0.0990 (0.006) | **0.0778 (0.004)**† | 0.0800 (0.004)† |
| ZDT1 d20 m2 | 3.01 (0.1) | 0.802 (0.2)† | **0.454 (0.07)**† | 0.823 (0.1) | 0.718 (0.1)† | 1.22 (0.2) | 1.48 (0.2) |
| ZDT2 d20 m2 | 3.81 (0.1) | 0.547 (0.1)† | **0.356 (0.08)**† | 1.11 (0.2) | 0.861 (0.09) | 1.88 (0.3) | 1.97 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.113 (0.004) | 0.0848 (0.005) | **0.0609 (0.006)**† | 0.0736 (0.007)† | 0.0777 (0.002) | 0.0751 (0.003) | 0.0917 (0.009) |
| DTLZ1 d15 m2 | 819 (70) | 635 (100) | 446 (50)† | 431 (40)† | 675 (111) | **384 (20)**† | 397 (40)† |
| DTLZ1 d15 m3 | 636 (50) | 461 (60) | 363 (10)† | 396 (30)† | 467 (40) | **331 (20)**† | 415 (80)† |
| DTLZ1 d20 m2 | 1119 (40) | 573 (50)† | 653 (90)† | 547 (40)† | 1271 (129) | **541 (30)**† | 604 (50)† |
| DTLZ1 d20 m3 | 1011 (40) | 594 (50) | 721 (80) | 712 (70) | 697 (90) | 447 (30)† | **425 (30)**† |
| DTLZ2 d15 m2 | 0.822 (0.08) | 0.772 (0.08) | 0.752 (0.07) | 0.688 (0.06) | 0.678 (0.05) | **0.363 (0.04)**† | 0.510 (0.07) |
| DTLZ2 d15 m3 | 0.679 (0.02) | 0.671 (0.03) | 0.675 (0.04) | 0.611 (0.04) | 0.530 (0.02) | 0.402 (0.006)† | **0.394 (0.007)**† |
| DTLZ2 d20 m2 | 0.982 (0.03) | 1.10 (0.1) | 1.11 (0.1) | 0.981 (0.08) | 0.879 (0.05) | **0.508 (0.1)**† | 0.840 (0.1) |
| DTLZ2 d20 m3 | 0.989 (0.02) | 1.09 (0.04) | 1.03 (0.07) | 0.845 (0.05) | 0.715 (0.03) | **0.419 (0.005)**† | 0.798 (0.3)† |
| DTLZ2 m5 | 0.634 (0.01) | 0.618 (0.02) | 0.617 (0.02) | 0.526 (0.03)† | **0.525 (0.02)**† | 0.988 (0.2) | 0.767 (0.1)† |
| DTLZ7 m5 | 0.207 (0.008) | 1.32e-4 (7e-5)† | **1.52e-5 (1e-5)**† | 6.96e-5 (4e-5)† | 0.0252 (0.005) | 1.98e-4 (1e-4)† | 4.33e-4 (2e-4) |
| Penicillin | 0.0925 (0.006)† | **0.0868 (0.01)**† | 0.126 (0.02)† | 0.106 (0.02)† | 0.102 (0.01)† | 0.129 (0.03) | 0.107 (0.007) |
| VehicleSafety | 0.199 (0.01) | 0.0500 (0.007) | **0.0206 (7e-4)**† | 0.0307 (0.003) | 0.0280 (0.002) | 0.0444 (0.005) | 0.0461 (0.006) |
| WFG6 m4 | 0.202 (0.02) | 0.158 (0.009)† | 0.158 (0.008)† | **0.141 (0.009)**† | 0.157 (0.009)† | 0.181 (0.009) | 0.162 (0.007)† |
| ZDT1 d20 m2 | 2.29 (0.1) | 0.940 (0.1)† | 0.976 (0.2)† | 1.27 (0.2)† | 1.08 (0.2)† | **0.860 (0.2)**† | 1.18 (0.1)† |
| ZDT2 d20 m2 | 3.44 (0.1) | 1.28 (0.2)† | 1.49 (0.4)† | 1.18 (0.1)† | **1.04 (0.08)**† | 1.37 (0.2)† | 1.72 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.119 (0.005) | 0.0865 (0.006) | 0.0789 (0.007) | 0.0779 (0.009) | 0.0830 (0.004) | **0.0578 (0.006)**† | 0.0953 (0.008) |
| DTLZ1 d15 m2 | 807 (60) | 435 (40) | 431 (60) | 445 (50) | 774 (70) | **265 (10)**† | 277 (9)† |
| DTLZ1 d15 m3 | 576 (40) | 377 (20) | 325 (30) | 411 (30) | 552 (30) | **227 (6)**† | 239 (10)† |
| DTLZ1 d20 m2 | 1202 (60) | 656 (50) | 747 (80) | 589 (60) | 1153 (80) | **368 (20)**† | 389 (30)† |
| DTLZ1 d20 m3 | 830 (40) | 481 (30) | 588 (60) | 661 (50) | 809 (30) | **338 (10)**† | 346 (20)† |
| DTLZ2 d15 m2 | 0.902 (0.04) | 0.965 (0.06) | 1.08 (0.08) | 1.20 (0.1) | 0.956 (0.07) | 0.644 (0.2)† | **0.604 (0.09)**† |
| DTLZ2 d15 m3 | 0.822 (0.02) | 0.770 (0.03) | 0.774 (0.05) | 0.723 (0.05)† | 0.593 (0.03)† | 0.759 (0.08)† | **0.502 (0.06)**† |
| DTLZ2 d20 m2 | 1.24 (0.06) | 1.73 (0.2) | 1.52 (0.1) | 1.31 (0.1) | 1.26 (0.1) | **0.595 (0.1)**† | 1.09 (0.1) |
| DTLZ2 d20 m3 | 1.08 (0.03) | 1.11 (0.04) | 1.19 (0.05) | 0.919 (0.05) | **0.752 (0.03)**† | 1.10 (0.08) | 0.888 (0.1)† |
| DTLZ2 m5 | 0.659 (0.01) | 0.681 (0.03) | 0.689 (0.03) | 0.611 (0.04)† | **0.583 (0.03)**† | 1.08 (0.2) | 1.08 (0.2) |
| DTLZ7 m5 | 0.200 (0.009) | 1.11e-4 (5e-5) | 2.70e-4 (2e-4) | 8.43e-5 (8e-5) | 0.0188 (0.005) | **0 (0)**† | 6.38e-4 (4e-4) |
| Penicillin | **0.0928 (0.008)**† | 0.124 (0.04)† | 0.195 (0.05)† | 0.0936 (0.007)† | 0.121 (0.03)† | 0.0967 (0.004)† | 0.122 (0.02)† |
| VehicleSafety | 0.199 (0.01) | 0.0225 (0.004)† | **0.0123 (0.002)**† | 0.0255 (0.007) | 0.0159 (0.002)† | 0.0196 (0.002) | 0.0196 (0.002) |
| WFG6 m4 | 0.140 (0.003) | 0.147 (0.004) | 0.148 (0.005) | 0.130 (0.004) | 0.140 (0.005) | 0.0901 (0.007)† | **0.0788 (0.005)**† |
| ZDT1 d20 m2 | 2.69 (0.3) | 1.28 (0.3)† | 1.16 (0.3)† | 1.93 (0.4) | 2.65 (0.07) | **0.590 (0.1)**† | 0.922 (0.2)† |
| ZDT2 d20 m2 | 3.54 (0.2) | 0.744 (0.1)† | **0.592 (0.1)**† | 0.685 (0.1)† | 2.27 (0.2) | 1.12 (0.2) | 1.30 (0.2) |

Table 26: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.4$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.148 (0.007) | 0.151 (0.01) | 0.152 (0.02)† | 0.144 (0.01)† | 0.160 (0.02) | **0.120 (0.02)**† | 0.132 (0.007) |
| DTLZ1 d15 m2 | 904 (60) | 525 (70)† | 735 (104) | 811 (100) | 826 (90) | **426 (40)**† | 429 (40)† |
| DTLZ1 d15 m3 | 618 (20) | 464 (40) | 481 (40) | 510 (50) | 375 (50)† | **293 (10)**† | 336 (10) |
| DTLZ1 d20 m2 | 1259 (60) | 912 (162) | 1153 (126) | 1210 (90) | 1135 (139) | 488 (30)† | **473 (20)**† |
| DTLZ1 d20 m3 | 959 (50) | 807 (90) | 793 (40) | 875 (80) | 891 (90) | **434 (30)**† | 439 (6)† |
| DTLZ2 d15 m2 | **0.912 (0.03)**† | 1.91 (0.3) | 1.72 (0.1) | 1.72 (0.1) | 1.33 (0.2)† | 1.10 (0.1) | 1.47 (0.3) |
| DTLZ2 d15 m3 | 0.839 (0.03)† | 0.858 (0.07)† | 0.997 (0.08) | 1.00 (0.08) | 0.722 (0.06)† | 0.898 (0.1)† | **0.700 (0.08)**† |
| DTLZ2 d20 m2 | **1.22 (0.02)**† | 2.09 (0.2) | 2.30 (0.3) | 2.04 (0.1) | 2.66 (0.4) | 1.77 (0.3) | 1.56 (0.3)† |
| DTLZ2 d20 m3 | 1.19 (0.03)† | 1.12 (0.1)† | 1.16 (0.1)† | 1.08 (0.09)† | **0.963 (0.08)**† | 1.08 (0.08)† | 1.17 (0.09)† |
| DTLZ2 m5 | 0.717 (0.01) | 0.592 (0.03)† | 0.654 (0.03) | **0.547 (0.02)**† | 0.588 (0.04)† | 0.696 (0.04) | 0.668 (0.05)† |
| DTLZ7 m5 | 0.211 (0.01) | 0.0534 (0.01) | 0.0766 (0.03) | 0.0987 (0.02) | 0.0420 (0.01) | **0.00684 (0.003)**† | 0.0361 (0.02)† |
| Penicillin | **0.110 (0.008)**† | 0.126 (0.01)† | 0.147 (0.02) | 0.130 (0.01)† | 0.135 (0.02)† | 0.165 (0.02) | 0.145 (0.02) |
| VehicleSafety | 0.277 (0.02) | 0.0925 (0.03)† | **0.0375 (0.01)**† | 0.0977 (0.03)† | 0.0810 (0.03) | 0.0651 (0.009) | 0.0627 (0.01)† |
| WFG6 m4 | 0.165 (0.005) | 0.164 (0.006) | 0.172 (0.005) | 0.177 (0.006) | 0.166 (0.01) | **0.101 (0.003)**† | 0.108 (0.007)† |
| ZDT1 d20 m2 | 3.05 (0.2)† | 3.03 (0.3)† | 3.05 (0.3)† | 3.79 (0.3) | 3.88 (0.2) | **2.36 (0.2)**† | 2.52 (0.3)† |
| ZDT2 d20 m2 | 3.99 (0.2) | 3.14 (0.3)† | **2.67 (0.2)**† | 3.30 (0.2)† | 3.11 (0.3)† | 2.91 (0.3)† | 2.78 (0.3)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.140 (0.006) | 0.106 (0.01)† | 0.113 (0.01) | 0.148 (0.03)† | 0.102 (0.01)† | **0.0821 (0.008)**† | 0.0956 (0.01)† |
| DTLZ1 d15 m2 | 888 (70) | 596 (128)† | 653 (80) | 595 (80)† | 620 (136)† | 406 (30)† | **401 (30)**† |
| DTLZ1 d15 m3 | 646 (20) | 374 (40)† | 399 (30) | 419 (20) | **284 (20)**† | 309 (8)† | 314 (30)† |
| DTLZ1 d20 m2 | 1224 (80) | 759 (70) | 957 (154) | 1087 (115) | 732 (100)† | 613 (100)† | **540 (40)**† |
| DTLZ1 d20 m3 | 938 (40) | 624 (80)† | 730 (70) | 765 (40) | 501 (50)† | 425 (20)† | **422 (20)**† |
| DTLZ2 d15 m2 | 1.01 (0.06)† | 1.30 (0.2)† | 1.71 (0.3)† | 1.37 (0.2)† | 0.991 (0.1)† | 1.63 (0.3) | **0.929 (0.1)**† |
| DTLZ2 d15 m3 | 0.843 (0.02) | 0.812 (0.04) | 0.739 (0.05)† | 0.811 (0.05)† | **0.681 (0.03)**† | 0.955 (0.1) | 0.745 (0.09)† |
| DTLZ2 d20 m2 | 1.32 (0.05)† | 2.03 (0.3) | 2.15 (0.2) | 1.96 (0.2) | **1.08 (0.1)**† | 1.67 (0.4)† | 1.66 (0.3)† |
| DTLZ2 d20 m3 | 1.18 (0.02) | 1.13 (0.06) | 1.05 (0.05) | 0.992 (0.07)† | **0.869 (0.05)**† | 1.08 (0.1) | 1.13 (0.09) |
| DTLZ2 m5 | 0.679 (0.02) | 0.629 (0.03)† | 0.647 (0.03)† | 0.618 (0.04)† | **0.580 (0.02)**† | 0.757 (0.03) | 0.657 (0.04)† |
| DTLZ7 m5 | 0.206 (0.01) | 0.0471 (0.01) | 0.0832 (0.02) | 0.0768 (0.02) | 0.0388 (0.01) | **6.37e-4 (4e-4)**† | 0.0259 (0.01) |
| Penicillin | 0.130 (0.02) | 0.121 (0.008) | 0.117 (0.009) | **0.0900 (0.005)**† | 0.139 (0.02) | 0.127 (0.01) | 0.114 (0.004) |
| VehicleSafety | 0.214 (0.01) | 0.0362 (0.01)† | 0.0374 (0.02)† | 0.0636 (0.02)† | 0.0445 (0.02)† | 0.0473 (0.006)† | **0.0303 (0.005)**† |
| WFG6 m4 | 0.165 (0.007) | 0.169 (0.005) | 0.176 (0.03) | 0.159 (0.009) | 0.128 (0.009) | 0.0967 (0.005)† | **0.0857 (0.005)**† |
| ZDT1 d20 m2 | 2.88 (0.1) | 2.37 (0.4) | **1.27 (0.1)**† | 2.20 (0.4)† | 1.41 (0.2)† | 2.81 (0.3) | 2.51 (0.4) |
| ZDT2 d20 m2 | 3.79 (0.09) | 2.18 (0.3) | **1.14 (0.1)**† | 2.01 (0.2) | 1.74 (0.2)† | 2.85 (0.2) | 3.26 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.126 (0.004)† | 0.137 (0.009) | **0.101 (0.01)**† | 0.140 (0.02)† | 0.130 (0.01) | 0.111 (0.01)† | 0.128 (0.01)† |
| DTLZ1 d15 m2 | 852 (50) | 495 (70)† | 649 (100)† | 861 (153) | 864 (70) | **457 (30)**† | 556 (40)† |
| DTLZ1 d15 m3 | 617 (40) | 509 (70) | 655 (80) | 654 (70) | 767 (70) | 334 (20)† | **333 (20)**† |
| DTLZ1 d20 m2 | 1138 (60) | 1018 (146) | 976 (136) | 1174 (146) | 1521 (149) | 584 (50)† | **565 (40)**† |
| DTLZ1 d20 m3 | 934 (40) | 766 (115)† | 824 (90) | 892 (70) | 1064 (122) | **433 (30)**† | 437 (20)† |
| DTLZ2 d15 m2 | 0.779 (0.04) | 0.714 (0.07) | 0.759 (0.04) | 0.844 (0.1) | 0.650 (0.03) | **0.379 (0.04)**† | 0.674 (0.1)† |
| DTLZ2 d15 m3 | 0.823 (0.04) | 0.729 (0.04)† | 0.739 (0.03)† | **0.663 (0.04)**† | 0.735 (0.05)† | 1.91 (0.4) | 1.08 (0.3)† |
| DTLZ2 d20 m2 | 1.04 (0.04)† | 0.841 (0.06)† | 1.03 (0.1)† | 1.00 (0.1)† | 0.857 (0.04)† | **0.811 (0.2)**† | 1.03 (0.2)† |
| DTLZ2 d20 m3 | 1.09 (0.03) | 0.899 (0.05)† | 0.920 (0.04)† | 0.871 (0.06)† | **0.869 (0.04)**† | 1.39 (0.3) | 1.79 (0.4) |
| DTLZ2 m5 | 0.716 (0.01) | 0.593 (0.03)† | 0.601 (0.02)† | 0.625 (0.02)† | **0.578 (0.02)**† | 0.769 (0.03) | 0.902 (0.07) |
| DTLZ7 m5 | 0.209 (0.008) | 0.00158 (0.001)† | 1.88e-4 (7e-5)† | 2.11e-4 (9e-5)† | 0.0491 (0.009) | **1.35e-4 (7e-5)**† | 0.00567 (0.004) |
| Penicillin | 0.120 (0.03)† | 0.191 (0.07)† | 0.126 (0.02)† | 0.101 (0.009)† | **0.0984 (0.005)**† | 0.111 (0.01)† | 0.107 (0.02)† |
| VehicleSafety | 0.256 (0.02) | 0.0575 (0.008) | **0.0306 (0.002)**† | 0.0341 (0.001)† | 0.0482 (0.009) | 0.0582 (0.01) | 0.0519 (0.005) |
| WFG6 m4 | 0.275 (0.02) | 0.298 (0.01) | 0.315 (0.04) | 0.239 (0.02)† | 0.295 (0.03)† | 0.299 (0.03) | **0.223 (0.02)**† |
| ZDT1 d20 m2 | 3.07 (0.2) | 2.79 (0.3) | **1.82 (0.3)**† | 2.91 (0.3) | 3.42 (0.3) | 2.25 (0.3)† | 2.72 (0.2) |
| ZDT2 d20 m2 | 3.70 (0.1) | **2.44 (0.4)**† | 2.86 (0.3)† | 2.52 (0.2)† | 2.92 (0.2)† | 2.50 (0.2)† | 2.81 (0.2)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.144 (0.006)† | 0.187 (0.03)† | 0.178 (0.02)† | **0.132 (0.02)**† | 0.153 (0.007) | 0.158 (0.03)† | 0.182 (0.03)† |
| DTLZ1 d15 m2 | 829 (40) | 691 (80) | 764 (100) | 792 (70) | 863 (40) | 420 (60)† | **389 (30)**† |
| DTLZ1 d15 m3 | 590 (40) | 544 (50) | 475 (30) | 631 (90) | 687 (40) | **272 (10)**† | 326 (30)† |
| DTLZ1 d20 m2 | 1308 (60) | 1094 (109) | 1204 (100) | 1289 (135) | 1231 (70) | **504 (40)**† | 550 (70)† |
| DTLZ1 d20 m3 | 888 (40) | 852 (40) | 833 (100) | 966 (70) | 984 (50) | 429 (30)† | **405 (30)**† |
| DTLZ2 d15 m2 | 0.979 (0.07)† | 0.973 (0.09)† | 0.919 (0.09)† | 0.968 (0.1)† | **0.909 (0.08)**† | 1.53 (0.2) | 1.37 (0.3)† |
| DTLZ2 d15 m3 | 0.850 (0.02) | 0.821 (0.05) | 0.830 (0.05) | 0.753 (0.03)† | **0.674 (0.01)**† | 1.40 (0.3) | 1.17 (0.2) |
| DTLZ2 d20 m2 | 1.35 (0.04) | 1.44 (0.2)† | 1.35 (0.1)† | 1.48 (0.2)† | **1.24 (0.1)**† | 1.72 (0.1) | 2.66 (0.5) |
| DTLZ2 d20 m3 | 1.16 (0.06) | 1.06 (0.05) | 1.17 (0.06) | 0.885 (0.06)† | **0.882 (0.02)**† | 1.48 (0.07) | 1.54 (0.05) |
| DTLZ2 m5 | 0.713 (0.01) | 0.689 (0.03)† | 0.699 (0.03)† | **0.601 (0.04)**† | 0.692 (0.04)† | 1.24 (0.2) | 1.47 (0.2) |
| DTLZ7 m5 | 0.210 (0.01) | 0.0263 (0.006) | 0.0494 (0.01) | 0.0934 (0.01) | 0.0726 (0.01) | **4.82e-5 (3e-5)**† | 0.00762 (0.005) |
| Penicillin | **0.108 (0.01)**† | 0.226 (0.06)† | 0.265 (0.06)† | 0.144 (0.01) | 0.115 (0.007)† | 0.117 (0.02)† | 0.135 (0.01) |
| VehicleSafety | 0.249 (0.02) | 0.131 (0.04)† | 0.0823 (0.02)† | **0.0458 (0.01)**† | 0.140 (0.02) | 0.0492 (0.01)† | 0.0917 (0.03)† |
| WFG6 m4 | 0.151 (0.003) | 0.165 (0.008) | 0.180 (0.009) | 0.167 (0.007) | 0.150 (0.008) | **0.113 (0.007)**† | **0.113 (0.007)**† |
| ZDT1 d20 m2 | 2.78 (0.3) | 2.51 (0.4)† | 2.88 (0.4) | 3.63 (0.3) | 2.78 (0.3) | **1.63 (0.3)**† | 2.73 (0.5)† |
| ZDT2 d20 m2 | 3.71 (0.2) | 3.52 (0.3) | 3.11 (0.4)† | 3.32 (0.2) | 3.81 (0.08) | **1.97 (0.4)**† | 2.82 (0.4)† |

Table 27: Terminal nIGD$^+$ at $\eta = 0.05$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.129 (0.002) | 0.0584 (0.002)† | **0.0577 (0.003)**† | 0.0878 (0.006) | 0.103 (0.002) | 0.0840 (0.004) | 0.108 (0.004) |
| DTLZ1 d15 m2 | 922 (70) | 299 (10)† | 329 (30)† | 328 (30)† | 321 (20)† | **279 (20)**† | 353 (30) |
| DTLZ1 d15 m3 | 692 (40) | 293 (10) | **249 (9)**† | 253 (20)† | 349 (30) | 265 (20)† | 260 (10)† |
| DTLZ1 d20 m2 | 1256 (80) | **399 (20)**† | 404 (30)† | 402 (20)† | 469 (30)† | 460 (10) | 494 (30) |
| DTLZ1 d20 m3 | 993 (50) | 329 (10)† | 371 (20)† | **325 (10)**† | 403 (20) | 403 (30) | 432 (20) |
| DTLZ2 d15 m2 | 0.851 (0.03) | **0.219 (0.01)**† | 0.249 (0.02)† | 0.287 (0.04)† | 1.19 (0.09) | 0.331 (0.01) | 0.725 (0.09) |
| DTLZ2 d15 m3 | 0.887 (0.02) | 0.631 (0.09)† | 0.557 (0.09)† | 0.902 (0.1) | 1.07 (0.03) | **0.550 (0.05)**† | 0.849 (0.05) |
| DTLZ2 d20 m2 | 1.21 (0.04) | 0.395 (0.05)† | 0.466 (0.1)† | 0.445 (0.04)† | 1.53 (0.06) | **0.339 (0.004)**† | 1.24 (0.06) |
| DTLZ2 d20 m3 | 1.23 (0.03)† | **0.845 (0.2)**† | 0.993 (0.1)† | 1.18 (0.1) | 1.42 (0.07) | 1.08 (0.1)† | 1.24 (0.02) |
| DTLZ2 m5 | 0.778 (0.02) | 0.645 (0.08)† | 0.603 (0.07)† | 0.800 (0.05) | 1.05 (0.1) | 0.714 (0.05) | **0.572 (0.05)**† |
| DTLZ7 m5 | 0.260 (0.005) | 0.0832 (0.004) | 0.0583 (0.005)† | **0.0409 (0.006)**† | 0.134 (0.004) | 0.0644 (0.007) | 0.0733 (0.008) |
| Penicillin | 0.118 (0.01)† | 0.130 (0.01) | 0.128 (0.01)† | 0.153 (0.02) | 0.114 (0.006)† | **0.105 (0.006)**† | 0.108 (0.004)† |
| VehicleSafety | 0.191 (0.009) | 0.0490 (0.003) | **0.0312 (0.002)**† | 0.0421 (0.003) | 0.0526 (0.003) | 0.0605 (0.004) | 0.0638 (0.005) |
| WFG6 m4 | 0.150 (0.001) | **0.118 (0.005)**† | 0.126 (0.004)† | 0.153 (0.002) | 0.151 (0.003) | 0.147 (0.004) | 0.133 (0.004)† |
| ZDT1 d20 m2 | 2.02 (0.09) | **0.104 (0.02)**† | 0.154 (0.02)† | 0.151 (0.03)† | 0.376 (0.01) | 0.461 (0.06) | 0.408 (0.08) |
| ZDT2 d20 m2 | 3.24 (0.09) | **0.0815 (0.02)**† | 0.0950 (0.01)† | 0.118 (0.03)† | 0.754 (0.06) | 0.756 (0.2) | 0.608 (0.1) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.128 (0.002) | **0.0575 (0.001)**† | 0.0578 (0.002)† | 0.0823 (0.004) | 0.102 (0.001) | 0.0881 (0.005) | 0.107 (0.005) |
| DTLZ1 d15 m2 | 879 (50) | 304 (20)† | 316 (20)† | 376 (90)† | 381 (30) | 294 (20)† | **293 (20)**† |
| DTLZ1 d15 m3 | 711 (40) | 301 (30) | 264 (10)† | **234 (8)**† | 313 (30) | 335 (40) | 286 (20) |
| DTLZ1 d20 m2 | 1209 (70) | **379 (20)**† | 394 (20)† | 419 (30)† | 433 (20)† | 473 (60)† | 403 (20)† |
| DTLZ1 d20 m3 | 1012 (50) | 364 (30)† | **337 (20)**† | 357 (30)† | 490 (40) | 437 (40) | 451 (40) |
| DTLZ2 d15 m2 | 0.940 (0.06) | 0.310 (0.05)† | **0.258 (0.03)**† | 0.392 (0.06)† | 1.17 (0.08) | 0.347 (0.01) | 0.942 (0.05) |
| DTLZ2 d15 m3 | 0.874 (0.02) | 0.732 (0.2)† | **0.579 (0.09)**† | 0.683 (0.1)† | 1.12 (0.05) | 0.622 (0.05)† | 0.942 (0.04) |
| DTLZ2 d20 m2 | 1.27 (0.06) | 0.383 (0.03)† | 0.353 (0.05)† | 0.470 (0.06)† | 1.54 (0.08) | **0.334 (0.005)**† | 1.26 (0.05) |
| DTLZ2 d20 m3 | 1.28 (0.05) | 1.14 (0.2)† | 1.11 (0.2)† | 1.11 (0.1)† | 1.43 (0.06) | **0.986 (0.1)**† | 1.35 (0.06) |
| DTLZ2 m5 | 0.783 (0.02) | 0.689 (0.07)† | 0.738 (0.2)† | 0.791 (0.06) | 1.13 (0.2) | 0.737 (0.06) | **0.503 (0.02)**† |
| DTLZ7 m5 | 0.259 (0.005) | 0.0634 (0.004) | 0.0635 (0.006) | **0.0419 (0.008)**† | 0.139 (0.004) | 0.0771 (0.01) | 0.0841 (0.007) |
| Penicillin | 0.126 (0.01)† | 0.124 (0.009)† | 0.134 (0.01)† | 0.209 (0.05) | 0.119 (0.009)† | **0.112 (0.007)**† | 0.113 (0.006)† |
| VehicleSafety | 0.190 (0.009) | 0.0511 (0.006) | **0.0315 (0.001)**† | 0.0427 (0.004) | 0.0495 (0.003) | 0.0660 (0.009) | 0.0836 (0.008) |
| WFG6 m4 | 0.147 (0.002) | **0.110 (0.003)**† | 0.118 (0.003) | 0.153 (0.003) | 0.149 (0.005) | 0.146 (0.004) | 0.128 (0.004) |
| ZDT1 d20 m2 | 2.03 (0.09) | **0.0995 (0.01)**† | 0.106 (0.01)† | 0.117 (0.02)† | 0.408 (0.02) | 0.445 (0.1) | 0.529 (0.08) |
| ZDT2 d20 m2 | 3.25 (0.09) | **0.0585 (0.005)**† | 0.0947 (0.01)† | 0.234 (0.03) | 0.806 (0.06) | 0.598 (0.09) | 0.836 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.130 (0.002) | 0.0656 (0.002) | **0.0584 (0.002)**† | 0.0849 (0.003) | 0.109 (0.001) | 0.0857 (0.003) | 0.101 (0.004) |
| DTLZ1 d15 m2 | 961 (50) | 439 (70)† | 364 (20)† | **340 (20)**† | 441 (50) | 367 (20)† | 373 (20)† |
| DTLZ1 d15 m3 | 687 (50) | 362 (30)† | 351 (20)† | **305 (20)**† | 376 (20) | 383 (40)† | 374 (20) |
| DTLZ1 d20 m2 | 1255 (40) | 505 (20)† | **472 (40)**† | 565 (100)† | 692 (123)† | 495 (30)† | 525 (30)† |
| DTLZ1 d20 m3 | 1113 (40) | 442 (30)† | 471 (40)† | **405 (30)**† | 651 (30) | 496 (30) | 547 (20) |
| DTLZ2 d15 m2 | 0.929 (0.05) | 0.290 (0.03)† | **0.246 (0.02)**† | 0.377 (0.06)† | 1.25 (0.06) | 0.371 (0.03) | 1.07 (0.1) |
| DTLZ2 d15 m3 | 0.893 (0.04) | 0.659 (0.1)† | 0.634 (0.1)† | 0.863 (0.06) | 1.00 (0.04) | **0.564 (0.09)**† | 0.754 (0.1)† |
| DTLZ2 d20 m2 | 1.18 (0.06) | 1.01 (0.1) | 0.592 (0.1)† | 0.676 (0.1) | 1.62 (0.08) | **0.347 (0.005)**† | 1.62 (0.09) |
| DTLZ2 d20 m3 | 1.18 (0.03) | 1.08 (0.2)† | 1.17 (0.1)† | 1.23 (0.1) | 1.45 (0.07) | **0.774 (0.1)**† | 1.51 (0.04) |
| DTLZ2 m5 | 0.794 (0.02)† | 1.13 (0.08) | 1.40 (0.2) | 1.13 (0.09) | 1.29 (0.2) | 1.92 (0.3) | **0.631 (0.06)**† |
| DTLZ7 m5 | 0.267 (0.004) | **0.0111 (0.002)**† | 0.0551 (0.004) | 0.0390 (0.006) | 0.152 (0.004) | 0.0907 (0.01) | 0.0806 (0.01) |
| Penicillin | 0.179 (0.03)† | 0.235 (0.04) | 0.198 (0.04)† | 0.250 (0.08)† | 0.163 (0.01) | 0.135 (0.009)† | **0.119 (0.01)**† |
| VehicleSafety | 0.197 (0.01) | 0.0555 (0.002)† | **0.0518 (8e-4)**† | 0.0607 (0.003)† | 0.0608 (0.002) | 0.0815 (0.004) | 0.0839 (0.006) |
| WFG6 m4 | 0.150 (0.002)† | **0.146 (0.003)**† | 0.149 (0.006)† | 0.163 (0.005) | 0.178 (0.005) | 0.190 (0.006) | 0.159 (0.003) |
| ZDT1 d20 m2 | 2.01 (0.08) | 0.431 (0.04) | **0.201 (0.02)**† | 0.254 (0.04)† | 0.462 (0.02) | 0.370 (0.04) | 0.439 (0.04) |
| ZDT2 d20 m2 | 3.27 (0.09) | 0.595 (0.02) | **0.283 (0.008)**† | 0.382 (0.05)† | 0.873 (0.08) | 0.497 (0.07) | 0.941 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.130 (0.003) | 0.0790 (0.004) | **0.0683 (0.003)**† | 0.0950 (0.005) | 0.108 (0.002) | 0.0817 (0.005) | 0.117 (0.005) |
| DTLZ1 d15 m2 | 961 (40) | 281 (10)† | 266 (10)† | 284 (10)† | 323 (10) | **259 (7)**† | 268 (7)† |
| DTLZ1 d15 m3 | 786 (40) | 255 (20)† | 289 (40)† | 276 (30)† | 294 (20)† | 253 (20)† | **241 (10)**† |
| DTLZ1 d20 m2 | 1211 (100) | **348 (4)**† | 357 (6)† | 366 (20)† | 445 (70)† | 402 (20) | 395 (30)† |
| DTLZ1 d20 m3 | 980 (30) | 352 (20)† | **325 (20)**† | 342 (10)† | 435 (40) | 370 (20) | 352 (20)† |
| DTLZ2 d15 m2 | 0.896 (0.04) | 0.213 (0.03)† | **0.173 (0.01)**† | 0.357 (0.08) | 1.35 (0.08) | 0.323 (0.005) | 1.27 (0.03) |
| DTLZ2 d15 m3 | 0.928 (0.03) | 0.581 (0.09)† | **0.374 (0.03)**† | 0.846 (0.07) | 0.951 (0.05) | 0.440 (0.01)† | 0.865 (0.1) |
| DTLZ2 d20 m2 | 1.26 (0.05) | 0.552 (0.2)† | 0.729 (0.2)† | 0.602 (0.09) | 1.44 (0.1) | **0.324 (0.003)**† | 2.14 (0.3) |
| DTLZ2 d20 m3 | 1.26 (0.04) | 1.26 (0.2) | 0.796 (0.2)† | 1.26 (0.1) | 1.45 (0.09) | **0.489 (0.06)**† | 1.51 (0.07) |
| DTLZ2 m5 | 0.866 (0.02) | 1.57 (0.2) | 1.42 (0.3) | 1.26 (0.1) | 1.05 (0.1) | 1.48 (0.3) | **0.559 (0.04)**† |
| DTLZ7 m5 | 0.261 (0.004) | 0.0366 (0.008)† | 0.0464 (0.004)† | **0.0305 (0.009)**† | 0.140 (0.006) | 0.105 (0.005) | 0.0777 (0.006) |
| Penicillin | 0.138 (0.02)† | 0.193 (0.04)† | 0.156 (0.03)† | 0.127 (0.01)† | 0.136 (0.03)† | **0.115 (0.008)**† | 0.137 (0.01)† |
| VehicleSafety | 0.191 (0.009) | 0.0443 (0.003) | **0.0354 (0.003)**† | 0.0376 (0.002)† | 0.0528 (0.004) | 0.0695 (0.006) | 0.0547 (0.006) |
| WFG6 m4 | 0.150 (0.002) | 0.139 (0.005)† | **0.129 (0.005)**† | 0.157 (0.002) | 0.163 (0.003) | 0.147 (0.005) | **0.129 (0.004)**† |
| ZDT1 d20 m2 | 2.03 (0.09) | 0.151 (0.01) | **0.0961 (0.009)**† | 0.221 (0.02) | 0.342 (0.01) | 0.234 (0.05) | 0.343 (0.05) |
| ZDT2 d20 m2 | 3.28 (0.1) | 0.114 (0.04)† | **0.0664 (0.007)**† | 0.176 (0.04)† | 0.980 (0.07) | 0.361 (0.04) | 0.745 (0.1) |

Table 28: Terminal nIGD$^+$ at $\eta = 0.2$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.198 (0.01) | 0.149 (0.008)† | **0.130 (0.006)**† | 0.175 (0.02) | 0.151 (0.006) | 0.146 (0.009)† | 0.169 (0.01) |
| DTLZ1 d15 m2 | 1028 (90) | **412 (70)**† | 462 (50)† | 543 (40) | 500 (70)† | 497 (90)† | 469 (50)† |
| DTLZ1 d15 m3 | 833 (60) | 472 (60) | 419 (50)† | 565 (50) | 412 (40) | 501 (60) | **313 (20)**† |
| DTLZ1 d20 m2 | 1372 (80) | **537 (40)**† | 541 (40)† | 687 (70)† | 568 (40)† | 635 (30) | 664 (60)† |
| DTLZ1 d20 m3 | 1221 (60) | 724 (104)† | 679 (90)† | 759 (70) | 579 (50)† | 559 (40)† | **511 (40)**† |
| DTLZ2 d15 m2 | 1.12 (0.04) | 1.20 (0.1) | 1.51 (0.2) | 1.34 (0.09) | 0.881 (0.07) | **0.551 (0.1)**† | 0.775 (0.1)† |
| DTLZ2 d15 m3 | 1.13 (0.03)† | 1.21 (0.05) | 1.25 (0.1) | 1.12 (0.06)† | 0.899 (0.05)† | 0.909 (0.1)† | **0.855 (0.1)**† |
| DTLZ2 d20 m2 | 1.38 (0.04) | 1.64 (0.2) | 1.82 (0.2) | 1.43 (0.2) | 1.19 (0.2) | **0.684 (0.1)**† | 1.28 (0.2) |
| DTLZ2 d20 m3 | 1.46 (0.04)† | 1.56 (0.07)† | 1.51 (0.08) | 1.25 (0.08)† | 1.21 (0.08)† | **1.09 (0.2)**† | 1.27 (0.1)† |
| DTLZ2 m5 | 1.05 (0.03) | 1.01 (0.04) | 1.02 (0.05) | 0.838 (0.04)† | 0.749 (0.03)† | 1.02 (0.06) | **0.712 (0.06)**† |
| DTLZ7 m5 | 0.355 (0.01) | 0.120 (0.01)† | **0.0951 (0.008)**† | 0.129 (0.01) | 0.151 (0.007) | 0.104 (0.01)† | 0.105 (0.009)† |
| Penicillin | **0.127 (0.007)**† | 0.141 (0.007)† | 0.134 (0.007)† | 0.171 (0.01) | 0.176 (0.02) | 0.177 (0.01) | 0.146 (0.009)† |
| VehicleSafety | 0.239 (0.02) | 0.0803 (0.009)† | **0.0567 (0.007)**† | 0.0720 (0.009)† | 0.0842 (0.007) | 0.0761 (0.008)† | 0.106 (0.009) |
| WFG6 m4 | 0.278 (0.005) | 0.207 (0.009)† | 0.219 (0.007) | 0.201 (0.005)† | 0.217 (0.01)† | 0.202 (0.01)† | **0.194 (0.006)**† |
| ZDT1 d20 m2 | 3.00 (0.1) | **0.938 (0.2)**† | 1.33 (0.3)† | 1.63 (0.2) | 1.63 (0.2)† | 1.52 (0.2) | 1.88 (0.3) |
| ZDT2 d20 m2 | 4.02 (0.2) | **1.09 (0.2)**† | 1.15 (0.3)† | 2.04 (0.2) | 1.24 (0.1)† | 2.18 (0.3) | 1.90 (0.3)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.163 (0.01) | 0.127 (0.007)† | **0.123 (0.007)**† | 0.124 (0.005)† | 0.131 (0.005)† | 0.142 (0.01)† | 0.146 (0.005) |
| DTLZ1 d15 m2 | 1014 (114) | 483 (60)† | 523 (50)† | 494 (50)† | **434 (50)**† | 465 (40)† | 444 (40)† |
| DTLZ1 d15 m3 | 736 (80) | 487 (50) | 373 (30)† | 461 (50) | **311 (20)**† | 356 (10)† | 366 (40)† |
| DTLZ1 d20 m2 | 1531 (116) | 682 (111)† | 664 (101)† | **580 (70)**† | 701 (169)† | 621 (50)† | 716 (182)† |
| DTLZ1 d20 m3 | 1323 (100) | 550 (70)† | 540 (80)† | 687 (80)† | 642 (90)† | 610 (70)† | **486 (20)**† |
| DTLZ2 d15 m2 | 1.38 (0.08) | 1.36 (0.2) | 1.11 (0.2) | 0.977 (0.1)† | 1.19 (0.1) | **0.833 (0.3)**† | 0.991 (0.2)† |
| DTLZ2 d15 m3 | 1.09 (0.04)† | 1.25 (0.07) | 1.22 (0.1)† | 1.21 (0.08) | **0.985 (0.05)**† | 0.993 (0.1)† | 0.999 (0.1)† |
| DTLZ2 d20 m2 | 1.67 (0.08) | 1.86 (0.2) | 1.44 (0.1) | 1.31 (0.2) | 1.23 (0.09) | **0.672 (0.1)**† | 1.44 (0.2) |
| DTLZ2 d20 m3 | 1.49 (0.04) | 1.58 (0.07) | 1.78 (0.2) | 1.47 (0.1)† | **1.22 (0.07)**† | 1.63 (0.2)† | 1.62 (0.07) |
| DTLZ2 m5 | 0.963 (0.03)† | 1.07 (0.04) | 1.10 (0.06) | **0.860 (0.05)**† | 1.02 (0.1)† | 1.27 (0.2)† | 0.995 (0.1)† |
| DTLZ7 m5 | 0.318 (0.01) | 0.103 (0.009)† | **0.0983 (0.01)**† | 0.101 (0.01)† | 0.129 (0.006) | 0.103 (0.01)† | 0.104 (0.008)† |
| Penicillin | **0.128 (0.008)**† | 0.152 (0.02)† | 0.129 (0.006)† | 0.150 (0.02)† | 0.177 (0.02) | 0.178 (0.02) | 0.161 (0.01) |
| VehicleSafety | 0.236 (0.01) | 0.0902 (0.01)† | **0.0648 (0.006)**† | 0.0813 (0.01)† | 0.0832 (0.01)† | 0.109 (0.01) | 0.0802 (0.006)† |
| WFG6 m4 | 0.228 (0.008) | 0.195 (0.005)† | 0.182 (0.009)† | 0.185 (0.006)† | **0.178 (0.007)**† | 0.182 (0.01)† | 0.188 (0.007)† |
| ZDT1 d20 m2 | 3.61 (0.2) | 0.863 (0.2)† | **0.507 (0.07)**† | 0.956 (0.1) | 0.870 (0.2)† | 1.47 (0.3) | 1.64 (0.2) |
| ZDT2 d20 m2 | 4.60 (0.2) | 0.561 (0.1)† | **0.385 (0.08)**† | 1.15 (0.2) | 0.913 (0.09) | 2.13 (0.4) | 1.98 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.155 (0.006) | 0.170 (0.03)† | **0.112 (0.01)**† | 0.114 (0.007)† | 0.131 (0.02)† | 0.136 (0.01)† | 0.189 (0.04) |
| DTLZ1 d15 m2 | 924 (60) | 649 (100) | 470 (50)† | 497 (70)† | 692 (108)† | 453 (30)† | **431 (30)**† |
| DTLZ1 d15 m3 | 893 (50) | 649 (70) | 592 (70)† | 801 (80) | 763 (90) | **423 (40)**† | 524 (70)† |
| DTLZ1 d20 m2 | 1198 (60) | 669 (122)† | 685 (100)† | 715 (80)† | 1314 (122) | **588 (40)**† | 645 (40)† |
| DTLZ1 d20 m3 | 1181 (70) | 1086 (148) | 942 (70) | 1289 (148) | 914 (111)† | 654 (90)† | **624 (70)**† |
| DTLZ2 d15 m2 | 0.884 (0.07)† | 0.893 (0.08)† | 0.898 (0.1)† | 0.965 (0.1)† | 0.784 (0.09)† | **0.709 (0.1)**† | 0.768 (0.1)† |
| DTLZ2 d15 m3 | 0.833 (0.03) | 0.976 (0.08) | 0.896 (0.02) | 0.903 (0.07) | **0.720 (0.02)**† | 1.05 (0.1) | 0.763 (0.08)† |
| DTLZ2 d20 m2 | 1.06 (0.05)† | 1.27 (0.1)† | 1.21 (0.1)† | 1.08 (0.1)† | 0.988 (0.07)† | **0.878 (0.2)**† | 1.33 (0.2)† |
| DTLZ2 d20 m3 | 1.10 (0.03)† | 1.37 (0.05) | 1.42 (0.08) | 1.30 (0.09) | **0.986 (0.04)**† | 1.35 (0.2)† | 1.17 (0.4)† |
| DTLZ2 m5 | 0.894 (0.03)† | 1.10 (0.06) | 0.921 (0.03)† | 0.904 (0.09)† | **0.839 (0.02)**† | 1.55 (0.3) | 1.59 (0.1) |
| DTLZ7 m5 | 0.288 (0.005) | 0.0341 (0.01)† | 0.0305 (0.006)† | **0.0285 (0.006)**† | 0.176 (0.006) | 0.0641 (0.008) | 0.0678 (0.009) |
| Penicillin | **0.125 (0.02)**† | 0.192 (0.02) | 0.217 (0.04) | 0.146 (0.02)† | 0.186 (0.02) | 0.176 (0.04)† | 0.179 (0.02) |
| VehicleSafety | 0.240 (0.02) | 0.0938 (0.009) | **0.0708 (0.004)**† | 0.0796 (0.007)† | 0.0715 (0.004)† | 0.107 (0.02) | 0.0845 (0.007)† |
| WFG6 m4 | 0.272 (0.02) | 0.231 (0.008) | 0.231 (0.01)† | **0.201 (0.01)**† | 0.227 (0.005) | 0.306 (0.02) | 0.240 (0.01) |
| ZDT1 d20 m2 | 2.32 (0.1) | 0.959 (0.1)† | 0.982 (0.2)† | 1.29 (0.2)† | 1.16 (0.2)† | **0.893 (0.1)**† | 1.24 (0.1) |
| ZDT2 d20 m2 | 3.58 (0.1) | 1.28 (0.2)† | 1.54 (0.4)† | 1.18 (0.1)† | **1.13 (0.1)**† | 1.37 (0.2)† | 1.76 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.204 (0.01) | **0.164 (0.007)**† | 0.213 (0.03)† | 0.181 (0.02)† | 0.171 (0.008)† | 0.217 (0.02) | 0.211 (0.02) |
| DTLZ1 d15 m2 | 923 (70) | 583 (50) | 545 (60) | 611 (90) | 896 (70) | **313 (20)**† | 369 (30)† |
| DTLZ1 d15 m3 | 724 (80) | 559 (70) | 528 (90) | 924 (100) | 774 (80) | **315 (20)**† | 336 (30)† |
| DTLZ1 d20 m2 | 1318 (90) | 804 (113) | 909 (102) | 678 (90)† | 1446 (102) | 503 (50)† | **481 (40)**† |
| DTLZ1 d20 m3 | 1133 (90) | 820 (80) | 839 (100) | 1141 (100) | 1279 (106) | **486 (40)**† | 516 (40)† |
| DTLZ2 d15 m2 | 1.02 (0.04)† | 1.38 (0.06) | 1.48 (0.1) | 1.51 (0.1) | 1.11 (0.06)† | **0.922 (0.2)**† | 0.993 (0.1)† |
| DTLZ2 d15 m3 | 1.08 (0.02)† | 1.16 (0.06) | 1.27 (0.1) | 1.12 (0.04)† | 0.966 (0.05)† | 1.29 (0.1) | **0.952 (0.07)**† |
| DTLZ2 d20 m2 | 1.50 (0.08)† | 2.01 (0.2) | 1.79 (0.1) | 1.68 (0.08) | 1.49 (0.1)† | **1.20 (0.2)**† | 1.61 (0.1) |
| DTLZ2 d20 m3 | 1.47 (0.04) | 1.83 (0.2) | 1.78 (0.09) | 1.44 (0.05) | **1.16 (0.04)**† | 1.63 (0.07) | 1.32 (0.2)† |
| DTLZ2 m5 | 0.992 (0.02) | 1.15 (0.06) | 1.17 (0.04) | 1.11 (0.1)† | **0.894 (0.04)**† | 1.69 (0.2) | 1.80 (0.2) |
| DTLZ7 m5 | 0.341 (0.006) | 0.0987 (0.009)† | 0.125 (0.01) | 0.0937 (0.01)† | 0.148 (0.01) | **0.0800 (0.02)**† | 0.114 (0.02)† |
| Penicillin | 0.157 (0.01)† | 0.180 (0.03)† | 0.242 (0.04)† | 0.153 (0.009)† | 0.182 (0.03)† | **0.150 (0.007)**† | 0.164 (0.02)† |
| VehicleSafety | 0.241 (0.02) | 0.0973 (0.009) | **0.0653 (0.006)**† | 0.0762 (0.009)† | 0.0863 (0.007)† | 0.0785 (0.009)† | 0.0843 (0.009)† |
| WFG6 m4 | 0.252 (0.006) | 0.250 (0.005) | 0.237 (0.008) | 0.215 (0.007)† | 0.242 (0.007) | **0.200 (0.02)**† | 0.205 (0.01)† |
| ZDT1 d20 m2 | 2.95 (0.3) | 1.38 (0.3)† | 1.22 (0.3)† | 1.97 (0.4) | 2.71 (0.09) | **0.617 (0.1)**† | 0.960 (0.2)† |
| ZDT2 d20 m2 | 3.63 (0.2) | 0.799 (0.1)† | **0.617 (0.1)**† | 0.705 (0.1)† | 2.34 (0.3) | 1.13 (0.2) | 1.30 (0.2) |

Table 29: Terminal nIGD$^+$ at $\eta = 0.4$, VHGP estimator: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.341 (0.01) | 0.271 (0.01)† | **0.265 (0.01)**† | 0.280 (0.01)† | 0.274 (0.02)† | 0.274 (0.02)† | 0.275 (0.02)† |
| DTLZ1 d15 m2 | 1149 (60) | 538 (60)† | 807 (111)† | 976 (80) | 853 (102) | 587 (60)† | **508 (60)**† |
| DTLZ1 d15 m3 | 992 (109) | 661 (70) | 807 (70) | 628 (70)† | 750 (100) | 512 (80)† | **444 (40)**† |
| DTLZ1 d20 m2 | 1318 (60) | 912 (162)† | 1174 (123) | 1262 (119) | 1191 (148)† | **775 (109)**† | 812 (144)† |
| DTLZ1 d20 m3 | 1213 (50) | 1025 (90)† | 1187 (100) | 1118 (103)† | 1127 (143)† | **768 (109)**† | 800 (70)† |
| DTLZ2 d15 m2 | **1.26 (0.05)**† | 1.93 (0.3)† | 1.73 (0.1) | 1.72 (0.1) | 1.45 (0.1)† | 1.46 (0.3)† | 1.68 (0.3)† |
| DTLZ2 d15 m3 | 1.23 (0.03)† | 1.42 (0.2)† | 1.37 (0.1)† | 1.33 (0.07)† | **1.12 (0.1)**† | 1.50 (0.2)† | 1.31 (0.2)† |
| DTLZ2 d20 m2 | **1.54 (0.07)**† | 2.09 (0.2)† | 2.36 (0.3) | 2.11 (0.1) | 2.66 (0.4) | 2.00 (0.3)† | 1.82 (0.3)† |
| DTLZ2 d20 m3 | 1.72 (0.06)† | 1.74 (0.1)† | **1.58 (0.1)**† | 1.61 (0.1)† | 1.75 (0.1)† | 1.96 (0.2)† | 1.75 (0.2)† |
| DTLZ2 m5 | 1.27 (0.06) | 1.08 (0.04)† | 1.15 (0.06)† | 1.00 (0.05)† | **0.999 (0.07)**† | 1.34 (0.1) | 1.37 (0.1)† |
| DTLZ7 m5 | 0.571 (0.02) | 0.226 (0.02)† | 0.261 (0.03)† | 0.254 (0.02) | 0.260 (0.02) | **0.180 (0.02)**† | 0.208 (0.03)† |
| Penicillin | 0.238 (0.02) | 0.195 (0.02)† | 0.215 (0.02)† | 0.190 (0.02)† | **0.174 (0.02)**† | 0.226 (0.03)† | 0.221 (0.02)† |
| VehicleSafety | 0.425 (0.04) | 0.148 (0.03)† | **0.143 (0.02)**† | 0.208 (0.05)† | 0.170 (0.03)† | 0.162 (0.02)† | 0.144 (0.02)† |
| WFG6 m4 | 0.395 (0.01) | 0.307 (0.01)† | 0.320 (0.01)† | 0.314 (0.01)† | **0.292 (0.01)**† | 0.334 (0.01)† | 0.302 (0.02)† |
| ZDT1 d20 m2 | 3.27 (0.2)† | 3.12 (0.2)† | **3.08 (0.2)**† | 3.86 (0.2) | 4.09 (0.2) | 3.28 (0.2)† | 3.21 (0.3)† |
| ZDT2 d20 m2 | 4.39 (0.2) | 3.14 (0.3)† | **2.73 (0.2)**† | 3.48 (0.3)† | 3.13 (0.3)† | 3.13 (0.3)† | 3.01 (0.4)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.257 (0.02) | 0.217 (0.02)† | 0.205 (0.01)† | 0.249 (0.04)† | 0.211 (0.02)† | 0.202 (0.02)† | **0.196 (0.01)**† |
| DTLZ1 d15 m2 | 1131 (70) | 596 (128)† | 790 (90) | 632 (80) | 647 (134)† | 662 (90) | **411 (40)**† |
| DTLZ1 d15 m3 | 882 (70) | 576 (100)† | 534 (50)† | 553 (60)† | 486 (90)† | **444 (40)**† | 496 (40)† |
| DTLZ1 d20 m2 | 1367 (110) | 857 (105)† | 1009 (151)† | 1126 (111) | **772 (112)**† | 888 (119)† | 850 (155)† |
| DTLZ1 d20 m3 | 1321 (90) | 909 (104) | 1061 (50) | 1018 (128) | **547 (70)**† | 662 (70)† | 613 (60)† |
| DTLZ2 d15 m2 | 1.27 (0.1)† | 1.40 (0.1)† | 1.80 (0.3) | 1.44 (0.2)† | **1.19 (0.1)**† | 1.89 (0.3)† | 1.32 (0.3)† |
| DTLZ2 d15 m3 | **1.18 (0.04)**† | 1.31 (0.09)† | 1.22 (0.07)† | 1.32 (0.1)† | 1.32 (0.1)† | 1.81 (0.4)† | 1.24 (0.2)† |
| DTLZ2 d20 m2 | 1.60 (0.07) | 2.08 (0.3) | 2.23 (0.2) | 2.18 (0.2) | **1.21 (0.1)**† | 2.56 (0.5) | 2.37 (0.4) |
| DTLZ2 d20 m3 | 1.70 (0.07) | 1.69 (0.1)† | 1.49 (0.07)† | 1.53 (0.1)† | **1.41 (0.1)**† | 2.33 (0.4)† | 1.51 (0.1)† |
| DTLZ2 m5 | 1.16 (0.03)† | 1.13 (0.09)† | 1.05 (0.04)† | 1.10 (0.05)† | **1.03 (0.08)**† | 1.33 (0.2)† | 1.08 (0.07)† |
| DTLZ7 m5 | 0.451 (0.01) | 0.211 (0.02)† | 0.237 (0.02) | 0.207 (0.02)† | 0.208 (0.01) | 0.171 (0.02)† | **0.161 (0.02)**† |
| Penicillin | 0.228 (0.02) | 0.176 (0.01)† | 0.177 (0.01)† | **0.152 (0.009)**† | 0.202 (0.02) | 0.187 (0.02)† | 0.214 (0.01) |
| VehicleSafety | 0.302 (0.02) | **0.108 (0.01)**† | 0.143 (0.02)† | 0.124 (0.02)† | 0.130 (0.02)† | 0.134 (0.01)† | 0.141 (0.02)† |
| WFG6 m4 | 0.315 (0.01) | 0.277 (0.01) | 0.279 (0.03) | 0.257 (0.01) | 0.232 (0.01)† | 0.281 (0.02) | **0.221 (0.01)**† |
| ZDT1 d20 m2 | 3.55 (0.3) | 2.57 (0.4) | **1.29 (0.1)**† | 2.23 (0.4)† | 1.66 (0.3)† | 3.48 (0.4) | 2.85 (0.4) |
| ZDT2 d20 m2 | 4.33 (0.2) | 2.18 (0.3) | **1.28 (0.2)**† | 2.06 (0.2) | 1.74 (0.2)† | 3.32 (0.3) | 3.64 (0.3) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | **0.209 (0.01)**† | 0.225 (0.02)† | 0.216 (0.03)† | 0.231 (0.03)† | 0.246 (0.03)† | 0.244 (0.02)† | 0.263 (0.05)† |
| DTLZ1 d15 m2 | 1001 (60) | 601 (100)† | 721 (103)† | 933 (149)† | 1144 (90) | **566 (50)**† | 594 (40)† |
| DTLZ1 d15 m3 | 1004 (60) | 920 (118)† | 1074 (70) | 882 (70) | 1250 (110) | 752 (104)† | **608 (90)**† |
| DTLZ1 d20 m2 | 1387 (90) | 1084 (158)† | 1119 (171) | 1174 (146) | 1655 (119) | 729 (50)† | **640 (40)**† |
| DTLZ1 d20 m3 | 1420 (60) | 1035 (130)† | 1046 (100)† | 1356 (123) | 1546 (151) | 1023 (111)† | **959 (131)**† |
| DTLZ2 d15 m2 | 0.986 (0.04) | **0.852 (0.1)**† | 0.857 (0.04)† | 0.919 (0.1)† | 0.880 (0.06)† | 1.11 (0.2)† | 1.07 (0.1)† |
| DTLZ2 d15 m3 | 1.09 (0.04)† | 0.984 (0.07)† | 1.02 (0.04)† | **0.953 (0.05)**† | 1.05 (0.04)† | 2.52 (0.3) | 1.80 (0.2) |
| DTLZ2 d20 m2 | 1.27 (0.04) | **0.987 (0.05)**† | 1.12 (0.1)† | 1.20 (0.1)† | 1.09 (0.09)† | 1.45 (0.1) | 1.38 (0.3)† |
| DTLZ2 d20 m3 | 1.48 (0.06) | 1.32 (0.09)† | 1.22 (0.07)† | **1.16 (0.06)**† | 1.39 (0.08) | 1.84 (0.3) | 3.11 (0.3) |
| DTLZ2 m5 | 1.14 (0.03) | 1.04 (0.03)† | 1.03 (0.04)† | 1.03 (0.04)† | **1.00 (0.04)**† | 1.91 (0.05) | 1.81 (0.1) |
| DTLZ7 m5 | 0.318 (0.005) | 0.0425 (0.01)† | **0.0358 (0.007)**† | 0.0549 (0.01)† | 0.211 (0.008) | 0.0604 (0.008)† | 0.0996 (0.02) |
| Penicillin | **0.150 (0.03)**† | 0.286 (0.06) | 0.208 (0.03) | 0.179 (0.02)† | 0.178 (0.02) | 0.164 (0.02)† | 0.151 (0.03)† |
| VehicleSafety | 0.346 (0.03) | 0.135 (0.02)† | **0.104 (0.01)**† | 0.148 (0.02)† | 0.136 (0.01) | 0.138 (0.02)† | 0.133 (0.02)† |
| WFG6 m4 | 0.376 (0.02)† | 0.384 (0.03) | 0.419 (0.04)† | **0.315 (0.03)**† | 0.413 (0.04) | 0.419 (0.03)† | 0.340 (0.02)† |
| ZDT1 d20 m2 | 3.77 (0.2) | 2.79 (0.3)† | **2.08 (0.4)**† | 2.91 (0.3) | 3.42 (0.3) | 3.17 (0.3) | 3.31 (0.2) |
| ZDT2 d20 m2 | 3.98 (0.1) | **2.44 (0.4)**† | 2.86 (0.3)† | 2.52 (0.2)† | 2.92 (0.2)† | 2.62 (0.2)† | 2.81 (0.2)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.291 (0.01)† | 0.336 (0.05)† | 0.273 (0.02)† | 0.272 (0.02)† | **0.267 (0.009)**† | 0.295 (0.03)† | 0.308 (0.03)† |
| DTLZ1 d15 m2 | 1077 (60) | 840 (90) | 921 (116) | 1012 (70) | 1127 (90) | 463 (60)† | **404 (20)**† |
| DTLZ1 d15 m3 | 985 (105) | 847 (100) | 835 (112) | 910 (80) | 987 (50) | **466 (50)**† | 485 (40)† |
| DTLZ1 d20 m2 | 1542 (136) | 1337 (118) | 1414 (128) | 1440 (107) | 1678 (100) | **538 (40)**† | 630 (50)† |
| DTLZ1 d20 m3 | 1181 (111) | 1172 (50) | 1279 (108) | 1477 (60) | 1424 (90) | 640 (70)† | **573 (70)**† |
| DTLZ2 d15 m2 | **1.12 (0.06)**† | 1.15 (0.1)† | 1.17 (0.05)† | 1.25 (0.1)† | 1.14 (0.07)† | 1.83 (0.2) | 1.78 (0.3) |
| DTLZ2 d15 m3 | 1.19 (0.02) | 1.09 (0.03)† | 1.13 (0.07)† | 1.07 (0.04)† | **1.00 (0.04)**† | 2.27 (0.2) | 1.82 (0.2) |
| DTLZ2 d20 m2 | **1.56 (0.05)**† | 1.64 (0.2)† | 1.76 (0.2)† | **1.56 (0.2)**† | 1.64 (0.1)† | 2.41 (0.2) | 2.94 (0.4) |
| DTLZ2 d20 m3 | 1.55 (0.04)† | 1.46 (0.06)† | 1.63 (0.06)† | 1.46 (0.08)† | **1.43 (0.07)**† | 2.45 (0.1) | 2.32 (0.06) |
| DTLZ2 m5 | 1.10 (0.03)† | 1.13 (0.04)† | 1.14 (0.03)† | **1.03 (0.06)**† | 1.10 (0.04)† | 2.11 (0.2) | 2.26 (0.1) |
| DTLZ7 m5 | 0.484 (0.01) | 0.195 (0.009) | 0.273 (0.02) | 0.285 (0.01) | 0.272 (0.02) | **0.126 (0.02)**† | 0.181 (0.03)† |
| Penicillin | **0.199 (0.02)**† | 0.287 (0.04)† | 0.328 (0.04) | 0.213 (0.03)† | 0.229 (0.02)† | 0.203 (0.02)† | 0.202 (0.02)† |
| VehicleSafety | 0.370 (0.03) | 0.228 (0.04)† | 0.183 (0.03)† | 0.173 (0.02)† | 0.234 (0.02) | **0.149 (0.02)**† | 0.161 (0.03)† |
| WFG6 m4 | 0.324 (0.006) | 0.344 (0.009) | 0.323 (0.01) | 0.286 (0.01) | 0.294 (0.01) | **0.193 (0.01)**† | 0.204 (0.01)† |
| ZDT1 d20 m2 | 3.05 (0.3) | 2.69 (0.4)† | 2.93 (0.4)† | 3.83 (0.3) | 2.88 (0.3) | **1.91 (0.3)**† | 2.86 (0.5)† |
| ZDT2 d20 m2 | 4.04 (0.2) | 3.59 (0.3) | 3.32 (0.4)† | 3.43 (0.2) | 4.18 (0.2) | **2.19 (0.5)**† | 3.00 (0.4)† |

Table 30: Terminal $\log_{10}$ HV gap at $\eta = 0.05$, oracle values (11), which score every evaluated design: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.61 (0.01) | **1.56 (0.01)**† | 1.98 (0.03) | 2.00 (0.008) | 1.96 (0.03) | 1.94 (0.01) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.29 (0.02)† | 4.26 (0.01)† | 4.35 (0.03) | 4.35 (0.05) | 4.26 (0.009)† | **4.25 (0.02)**† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.30 (0.02)† | **6.29 (0.05)**† | 6.56 (0.04) | 6.64 (0.05) | 6.30 (0.03)† | **6.29 (0.03)**† |
| DTLZ1 d20 m2 | 5.19 (0.006) | **4.51 (0.008)**† | 4.56 (0.02) | 4.59 (0.02) | 4.66 (0.03) | 4.57 (0.02) | 4.55 (0.01) |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.74 (0.04)† | 6.80 (0.03) | 7.04 (0.04) | 7.09 (0.03) | 6.74 (0.01)† | **6.70 (0.02)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.671 (0.02)† | **-0.678 (0.02)**† | -0.560 (0.03) | -0.372 (0) | -0.599 (0.004) | -0.438 (0.03) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.178 (0.03)† | **-0.194 (0.02)**† | -0.124 (0.01) | -0.0932 (2e-4) | -0.190 (0.004)† | -0.103 (0.01) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.524 (0.03)† | -0.553 (0.03)† | -0.457 (0.02) | -0.372 (0) | **-0.585 (0.003)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | **-0.152 (0.02)**† | -0.122 (0.01)† | -0.100 (0.003)† | -0.0929 (0) | -0.128 (0.01)† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003) | 0.0452 (0.02)† | **0.0373 (0.02)**† | 0.127 (0.009) | 0.153 (0.002) | 0.124 (0.003) | 0.110 (0.002) |
| Penicillin | 5.68 (0.004) | 5.59 (0.02)† | 5.61 (0.03)† | 5.60 (0.02)† | 5.62 (0.02)† | 5.59 (0.02)† | **5.58 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | -0.0627 (0.02) | **-0.444 (0.03)**† | 0.384 (0.03) | 0.267 (0.04) | 0.340 (0.07) | 0.272 (0.03) |
| WFG6 m4 | 2.65 (0.003) | **2.13 (0.009)**† | 2.14 (0.02)† | 2.40 (0.02) | 2.32 (0.03) | 2.21 (0.007) | 2.16 (0.01)† |
| ZDT1 d20 m2 | 1.41 (0.02) | **-0.138 (0.2)**† | 0.219 (0.1)† | 0.495 (0.08) | 0.577 (0.05) | 0.718 (0.06) | 0.330 (0.2)† |
| ZDT2 d20 m2 | 1.62 (0.01) | **-0.816 (0.2)**† | -0.631 (0.1)† | -0.164 (0.4)† | 1.07 (0.05) | 1.06 (0.1) | 0.868 (0.2) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.61 (0.01)† | **1.57 (0.02)**† | 1.92 (0.03) | 2.02 (0.004) | 1.95 (0.01) | 1.94 (0.02) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.33 (0.03) | 4.31 (0.03) | 4.32 (0.02)† | 4.38 (0.03) | **4.24 (0.003)**† | 4.29 (0.02) |
| DTLZ1 d15 m3 | 7.50 (0.02) | **6.29 (0.02)**† | 6.31 (0.02)† | 6.61 (0.07) | 6.60 (0.06) | **6.29 (0.03)**† | 6.32 (0.05)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | **4.54 (0.02)**† | 4.55 (0.02)† | 4.63 (0.02) | 4.61 (0.03)† | 4.56 (0.01)† | 4.59 (0.01) |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.74 (0.04)† | **6.71 (0.03)**† | 7.01 (0.04) | 7.06 (0.05) | 6.77 (0.01)† | 6.84 (0.06)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.608 (0.03)† | **-0.657 (0.03)**† | -0.522 (0.04) | -0.372 (0) | -0.605 (0.005) | -0.372 (0) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | **-0.198 (0.03)**† | -0.193 (0.03)† | -0.148 (0.01)† | -0.0935 (3e-4) | -0.187 (0.002)† | -0.0930 (3e-5) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.525 (0.02) | -0.580 (0.03)† | -0.487 (0.03) | -0.372 (0) | **-0.589 (0.002)**† | -0.392 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.129 (0.02)† | -0.132 (0.02)† | -0.107 (0.007)† | -0.0929 (0) | **-0.147 (0.01)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003) | **0.0367 (0.03)**† | 0.0418 (0.02)† | 0.124 (0.01) | 0.154 (0.002) | 0.128 (0.007) | 0.107 (0.003)† |
| Penicillin | 5.68 (0.004) | 5.62 (0.01)† | 5.60 (0.02)† | 5.60 (0.02)† | 5.63 (0.02)† | 5.60 (0.02)† | **5.59 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | -0.0875 (0.03) | **-0.399 (0.03)**† | 0.358 (0.03) | 0.294 (0.03) | 0.423 (0.06) | 0.376 (0.04) |
| WFG6 m4 | 2.65 (0.003) | 2.13 (0.01)† | **2.11 (0.01)**† | 2.43 (0.02) | 2.33 (0.02) | 2.20 (0.006) | 2.16 (0.008) |
| ZDT1 d20 m2 | 1.41 (0.02) | **-0.0641 (0.2)**† | 0.0604 (0.2)† | 0.439 (0.06) | 0.689 (0.03) | 0.481 (0.2) | 0.668 (0.1) |
| ZDT2 d20 m2 | 1.62 (0.01) | **-1.09 (0.07)**† | -0.744 (0.06) | 0.671 (0.2) | 1.14 (0.03) | 0.928 (0.2) | 0.909 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.28 (0.009) | 1.80 (0.008) | **1.75 (0.008)**† | 1.90 (0.01) | 2.07 (0.008) | 2.02 (0.01) | 1.96 (0.01) |
| DTLZ1 d15 m2 | 5.09 (0.02) | 4.49 (0.02) | 4.43 (0.01)† | 4.47 (0.02)† | 4.46 (0.01) | **4.42 (0.005)**† | 4.45 (0.02)† |
| DTLZ1 d15 m3 | 7.52 (0.02) | 6.55 (0.02) | 6.54 (0.04)† | 6.70 (0.06) | 6.76 (0.04) | **6.48 (0.02)**† | 6.59 (0.04) |
| DTLZ1 d20 m2 | 5.20 (0.005) | 4.71 (0.01) | **4.67 (0.01)**† | **4.67 (0.02)**† | 4.76 (0.04)† | **4.67 (0.009)**† | 4.71 (0.02) |
| DTLZ1 d20 m3 | 7.75 (0.008) | 6.99 (0.04) | 6.90 (0.03)† | 7.04 (0.04) | 7.31 (0.06) | **6.84 (0.02)**† | 7.01 (0.03) |
| DTLZ2 d15 m2 | -0.372 (0) | -0.571 (0.03)† | **-0.647 (0.02)**† | -0.536 (0.03) | -0.372 (0) | -0.599 (0.004) | -0.395 (0.02) |
| DTLZ2 d15 m3 | -0.0967 (0.002) | -0.198 (0.02)† | **-0.201 (0.02)**† | -0.134 (0.01)† | -0.0973 (0.004) | -0.197 (0.004)† | -0.138 (0.02) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.412 (0.02) | -0.497 (0.03)† | -0.435 (0.02) | -0.372 (0) | **-0.576 (0.004)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.139 (0.02)† | -0.109 (0.009) | -0.0970 (0.003) | -0.0932 (2e-4) | **-0.182 (0.002)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.140 (0.002)† | **0.0864 (0.02)**† | 0.0967 (0.02)† | 0.133 (0.01) | 0.147 (0.003) | 0.115 (0.003)† | 0.0913 (0.005)† |
| Penicillin | 5.66 (0.005) | **5.47 (0.03)**† | 5.56 (0.02) | 5.55 (0.02)† | 5.60 (0.01) | 5.56 (0.02) | 5.56 (0.02) |
| VehicleSafety | 1.24 (0.01) | 0.783 (0.003) | **0.758 (8e-4)**† | 0.850 (0.008) | 0.811 (0.004) | 0.847 (0.01) | 0.855 (0.01) |
| WFG6 m4 | 2.65 (0.003) | 2.43 (0.003)† | **2.42 (0.004)**† | 2.53 (0.008) | 2.49 (0.008) | 2.45 (0.002) | 2.46 (0.005) |
| ZDT1 d20 m2 | 1.40 (0.02) | 0.653 (0.08) | **0.349 (0.06)**† | 0.639 (0.06) | 0.737 (0.03) | 0.549 (0.06) | 0.715 (0.06) |
| ZDT2 d20 m2 | 1.63 (0.01) | 1.12 (0.008) | **0.700 (0.03)**† | 0.877 (0.07)† | 1.14 (0.03) | 0.864 (0.07) | 1.13 (0.08) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 1.77 (0.01) | **1.61 (0.02)**† | 1.93 (0.03) | 2.00 (0.008) | 1.97 (0.02) | 1.89 (0.008) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.29 (0.02) | 4.27 (0.02) | 4.29 (0.02)† | 4.40 (0.03) | **4.24 (0.004)**† | 4.25 (0.007)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.33 (0.05) | 6.24 (0.01) | 6.58 (0.08) | 6.73 (0.08) | **6.16 (0.02)**† | 6.20 (0.02)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.52 (0.009)† | **4.51 (0.01)**† | 4.53 (0.01)† | 4.67 (0.06) | 4.53 (0.01)† | 4.55 (0.01) |
| DTLZ1 d20 m3 | 7.74 (0.009) | 6.72 (0.03)† | 6.68 (0.02)† | 6.97 (0.05) | 7.20 (0.05) | **6.66 (0.02)**† | 6.73 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.687 (0.03)† | **-0.750 (0.02)**† | -0.544 (0.03) | -0.372 (0) | -0.605 (0.002) | -0.372 (0) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.214 (0.03)† | **-0.272 (0.02)**† | -0.159 (0.01) | -0.0969 (0.003) | -0.193 (0.002) | -0.134 (0.02) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.566 (0.04)† | -0.557 (0.04)† | -0.437 (0.02) | -0.372 (0) | **-0.593 (0.003)**† | -0.372 (0) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.129 (0.02) | -0.179 (0.02)† | -0.111 (0.009) | -0.0929 (0) | **-0.187 (0.002)**† | -0.0929 (0) |
| DTLZ2 m5 | 0.147 (0.003)† | 0.102 (0.02)† | **0.0647 (0.03)**† | 0.146 (0.005) | 0.144 (0.005) | 0.131 (0.005)† | 0.0985 (0.008)† |
| Penicillin | 5.68 (0.004) | 5.60 (0.02)† | **5.57 (0.03)**† | 5.60 (0.02) | 5.64 (0.01) | 5.61 (0.02)† | **5.57 (0.02)**† |
| VehicleSafety | 1.21 (0.01) | -0.107 (0.02) | **-0.436 (0.04)**† | 0.369 (0.04) | 0.318 (0.03) | 0.406 (0.06) | 0.214 (0.03) |
| WFG6 m4 | 2.65 (0.003) | 2.13 (0.004)† | **2.12 (0.01)**† | 2.41 (0.03) | 2.34 (0.008) | 2.20 (0.009) | 2.14 (0.01)† |
| ZDT1 d20 m2 | 1.41 (0.02) | 0.354 (0.05)† | 0.0732 (0.1)† | 0.671 (0.05) | 0.567 (0.03) | **-0.0145 (0.2)**† | 0.438 (0.1)† |
| ZDT2 d20 m2 | 1.62 (0.01) | -0.824 (0.09)† | **-0.952 (0.07)**† | 0.115 (0.3)† | 1.22 (0.02) | 0.608 (0.1) | 1.00 (0.1) |

Table 31: Terminal $\log_{10}$ HV gap at $\eta = 0.2$, oracle values (11), which score every evaluated design: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 2.05 (0.01) | **1.94 (0.03)**† | 2.15 (0.03) | 1.98 (0.01)† | 1.95 (0.01)† | 2.02 (0.02)† |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.37 (0.07)† | 4.45 (0.05) | 4.47 (0.05) | 4.56 (0.08) | **4.31 (0.02)**† | 4.33 (0.02)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.94 (0.07) | 6.84 (0.08) | 7.11 (0.06) | 6.83 (0.09) | **6.44 (0.07)**† | 6.51 (0.07)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.64 (0.05)† | 4.75 (0.05) | 4.77 (0.05) | 4.75 (0.04) | **4.62 (0.03)**† | 4.65 (0.03)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 7.37 (0.08) | 7.44 (0.07) | 7.55 (0.03) | 7.11 (0.07) | **6.87 (0.04)**† | 6.93 (0.07)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.373 (1e-3) | -0.385 (0.01) | -0.389 (0.008) | -0.376 (0.003) | **-0.595 (0.004)**† | -0.495 (0.04)† |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.0963 (0.003) | -0.0991 (0.005)† | -0.108 (0.01)† | -0.0954 (0.001)† | **-0.147 (0.01)**† | **-0.147 (0.02)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.377 (0.004) | -0.376 (0.004) | -0.376 (0.004) | **-0.587 (0.003)**† | -0.436 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0943 (8e-4)† | -0.0933 (4e-4)† | **-0.126 (0.01)**† | -0.118 (0.01)† |
| DTLZ2 m5 | 0.147 (0.003) | 0.145 (0.004) | 0.144 (0.005) | 0.121 (0.008)† | **0.116 (0.008)**† | 0.137 (0.006)† | 0.122 (0.008)† |
| Penicillin | 5.68 (0.004) | 5.64 (0.008)† | 5.63 (0.01)† | 5.63 (0.03)† | 5.66 (0.01)† | **5.61 (0.02)**† | 5.63 (0.02)† |
| VehicleSafety | 1.21 (0.01) | 0.402 (0.04) | **0.104 (0.05)**† | 0.529 (0.04) | 0.216 (0.02)† | 0.383 (0.03) | 0.447 (0.07) |
| WFG6 m4 | 2.65 (0.003) | 2.49 (0.03) | 2.53 (0.03) | 2.59 (0.004) | 2.43 (0.03) | **2.20 (0.006)**† | 2.29 (0.02) |
| ZDT1 d20 m2 | 1.41 (0.02) | 0.967 (0.04)† | **0.955 (0.07)**† | 1.24 (0.03) | 1.08 (0.07)† | 1.09 (0.05)† | 1.12 (0.04) |
| ZDT2 d20 m2 | 1.62 (0.01) | **0.984 (0.2)**† | 1.00 (0.1)† | 1.29 (0.07)† | 1.17 (0.05)† | 1.40 (0.03) | 1.34 (0.05) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 2.01 (0.03)† | 1.96 (0.02)† | 2.08 (0.03) | 1.98 (0.02)† | **1.95 (0.009)**† | 2.01 (0.01) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.54 (0.05) | 4.53 (0.05) | 4.46 (0.05) | 4.44 (0.06)† | 4.33 (0.02)† | **4.32 (0.02)**† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 6.81 (0.1) | 6.69 (0.07) | 7.00 (0.06) | 6.49 (0.06)† | **6.39 (0.05)**† | 6.55 (0.08)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.73 (0.05)† | 4.75 (0.05) | 4.80 (0.05) | 4.72 (0.07)† | 4.61 (0.01)† | **4.60 (0.02)**† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 7.36 (0.06) | 7.30 (0.1) | 7.49 (0.07) | 7.11 (0.09)† | **6.90 (0.04)**† | 6.94 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.378 (0.004) | -0.395 (0.01) | -0.395 (0.01) | -0.373 (6e-4) | **-0.578 (0.02)**† | -0.439 (0.03) |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.0965 (0.002) | -0.0963 (0.002) | -0.104 (0.009) | -0.0975 (0.002) | **-0.161 (0.01)**† | -0.122 (0.01) |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.378 (0.004) | -0.380 (0.004) | -0.372 (0) | **-0.555 (0.02)**† | -0.416 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0)† | -0.0933 (4e-4)† | -0.0929 (0)† | -0.0938 (6e-4)† | -0.0947 (0.002)† | **-0.112 (0.01)**† | -0.0929 (0)† |
| DTLZ2 m5 | 0.147 (0.003) | 0.146 (0.005) | 0.145 (0.004) | 0.124 (0.006)† | **0.117 (0.007)**† | 0.140 (0.005) | 0.124 (0.008)† |
| Penicillin | 5.68 (0.004) | 5.65 (0.01)† | **5.60 (0.01)**† | 5.62 (0.02)† | 5.64 (0.01)† | **5.60 (0.02)**† | 5.63 (0.02)† |
| VehicleSafety | 1.21 (0.01) | 0.293 (0.04) | **0.0490 (0.03)**† | 0.363 (0.05) | 0.286 (0.05) | 0.429 (0.07) | 0.397 (0.03) |
| WFG6 m4 | 2.65 (0.003) | 2.48 (0.03) | 2.42 (0.05) | 2.58 (0.009) | 2.42 (0.04) | **2.21 (0.01)**† | 2.28 (0.02) |
| ZDT1 d20 m2 | 1.41 (0.02) | 0.961 (0.07) | **0.569 (0.08)**† | 0.977 (0.1) | 0.908 (0.08) | 0.990 (0.07) | 1.21 (0.04) |
| ZDT2 d20 m2 | 1.62 (0.01) | 0.634 (0.2)† | **0.356 (0.2)**† | 1.22 (0.06) | 1.13 (0.05) | 1.35 (0.05) | 1.44 (0.03) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.28 (0.009) | 1.99 (0.01) | **1.91 (0.02)**† | 1.95 (0.02)† | 2.10 (0.01) | 1.98 (0.007) | 2.00 (0.02) |
| DTLZ1 d15 m2 | 5.09 (0.02) | 4.69 (0.03) | 4.62 (0.05) | 4.56 (0.04) | 4.80 (0.09) | **4.43 (0.005)**† | **4.43 (0.01)**† |
| DTLZ1 d15 m3 | 7.52 (0.02) | 6.95 (0.08) | 7.04 (0.08) | 7.10 (0.07) | 7.07 (0.1) | **6.49 (0.02)**† | 6.60 (0.05) |
| DTLZ1 d20 m2 | 5.20 (0.005) | 4.83 (0.03) | 4.86 (0.05) | 4.82 (0.03) | 5.16 (0.04) | 4.71 (0.02)† | **4.70 (0.01)**† |
| DTLZ1 d20 m3 | 7.75 (0.008) | 7.41 (0.06) | 7.47 (0.04) | 7.49 (0.06) | 7.50 (0.08) | 6.96 (0.04)† | **6.89 (0.008)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.386 (0.006) | -0.393 (0.01) | -0.382 (0.007) | -0.382 (0.01) | **-0.594 (0.006)**† | -0.515 (0.03) |
| DTLZ2 d15 m3 | -0.0967 (0.002) | -0.0988 (0.003) | -0.101 (0.004) | -0.109 (0.008) | -0.109 (0.006) | -0.192 (0.003)† | **-0.195 (0.007)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.376 (0.004) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.578 (0.004)**† | -0.420 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0945 (0.001) | -0.0929 (0) | -0.100 (0.004) | **-0.178 (0.003)**† | **-0.178 (0.01)**† |
| DTLZ2 m5 | 0.140 (0.002) | 0.142 (0.004) | 0.142 (0.005) | 0.108 (0.01)† | 0.110 (0.007)† | 0.134 (0.004) | **0.101 (0.005)**† |
| Penicillin | 5.66 (0.005) | **5.49 (0.02)**† | 5.55 (0.02)† | 5.56 (0.03)† | 5.57 (0.02) | 5.55 (0.02)† | 5.56 (0.02)† |
| VehicleSafety | 1.24 (0.01) | 0.847 (0.01) | **0.782 (0.003)**† | 0.835 (0.01) | 0.814 (0.005) | 0.872 (0.02) | 0.850 (0.01) |
| WFG6 m4 | 2.65 (0.003) | 2.53 (0.02)† | 2.55 (0.01) | 2.58 (0.01) | 2.53 (0.02)† | **2.48 (0.01)**† | 2.50 (0.008)† |
| ZDT1 d20 m2 | 1.40 (0.02) | 1.04 (0.04) | 0.903 (0.05)† | 1.14 (0.05) | 1.05 (0.05)† | **0.855 (0.07)**† | 0.966 (0.07)† |
| ZDT2 d20 m2 | 1.63 (0.01) | 1.28 (0.03)† | **1.17 (0.09)**† | 1.25 (0.03)† | 1.21 (0.03)† | 1.19 (0.07)† | 1.29 (0.04)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 2.08 (0.01) | 2.05 (0.02)† | 2.10 (0.04) | 2.07 (0.01) | **1.98 (0.03)**† | 2.00 (0.02)† |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.63 (0.04) | 4.60 (0.05) | 4.58 (0.06) | 5.01 (0.03) | **4.24 (0.007)**† | 4.27 (0.01) |
| DTLZ1 d15 m3 | 7.50 (0.02) | 7.01 (0.07) | 7.01 (0.07) | 7.25 (0.05) | 7.45 (0.02) | **6.19 (0.02)**† | 6.24 (0.01) |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.90 (0.06) | 4.92 (0.04) | 4.90 (0.05) | 5.19 (0.006) | **4.53 (0.008)**† | 4.55 (0.009)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 7.42 (0.06) | 7.44 (0.04) | 7.60 (0.05) | 7.72 (0.01) | **6.69 (0.01)**† | 6.74 (0.03)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (2e-5) | -0.386 (0.009) | -0.374 (0.002) | -0.372 (0) | **-0.588 (0.003)**† | -0.518 (0.03)† |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.0988 (0.003) | -0.0990 (0.005) | -0.0970 (0.004) | -0.0987 (0.003) | -0.128 (0.01) | **-0.177 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.496 (0.03)**† | -0.392 (0.02) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.0934 (5e-4) | -0.0931 (1e-4) | -0.105 (0.009)† | **-0.131 (0.02)**† |
| DTLZ2 m5 | 0.147 (0.003) | 0.146 (0.004) | 0.148 (0.005) | 0.129 (0.01)† | 0.127 (0.008)† | 0.154 (0.002) | **0.121 (0.007)**† |
| Penicillin | 5.68 (0.004) | **5.59 (0.03)**† | 5.61 (0.02)† | 5.65 (0.01)† | 5.62 (0.02)† | 5.60 (0.02)† | 5.61 (0.02)† |
| VehicleSafety | 1.21 (0.01) | 0.450 (0.07) | **0.215 (0.05)**† | 0.501 (0.06) | 0.309 (0.06)† | 0.377 (0.04) | 0.408 (0.05) |
| WFG6 m4 | 2.65 (0.003) | 2.60 (0.01) | 2.60 (0.01) | 2.60 (0.008) | 2.62 (0.01) | **2.21 (0.02)**† | 2.23 (0.02)† |
| ZDT1 d20 m2 | 1.41 (0.02) | 1.00 (0.1) | 0.973 (0.1) | 1.20 (0.07) | 1.41 (0.02) | **0.609 (0.1)**† | 0.814 (0.2) |
| ZDT2 d20 m2 | 1.62 (0.01) | 0.995 (0.1)† | **0.791 (0.2)**† | 1.06 (0.07)† | 1.46 (0.04) | 1.18 (0.07) | 1.25 (0.04) |

Table 32: Terminal $\log_{10}$ HV gap at $\eta = 0.4$, oracle values (11), which score every evaluated design: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better. DTLZ7 $m$=5 omitted: no known maximal hypervolume.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 2.14 (0.03) | 2.14 (0.03) | 2.19 (0.02) | 2.15 (0.03) | **1.99 (0.01)**† | 2.11 (0.02) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.49 (0.07) | 4.68 (0.06) | 4.80 (0.04) | 4.87 (0.06) | **4.32 (0.02)**† | **4.32 (0.02)**† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 7.14 (0.08) | 7.28 (0.04) | 7.31 (0.04) | 7.09 (0.1) | **6.41 (0.06)**† | 6.48 (0.05)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.95 (0.06) | 5.07 (0.04) | 5.04 (0.04) | 5.07 (0.06) | 4.63 (0.02)† | **4.62 (0.02)**† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 7.62 (0.03) | 7.69 (0.01) | 7.66 (0.03) | 7.58 (0.08) | **6.91 (0.04)**† | 6.92 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (3e-4) | -0.374 (0.001) | -0.374 (0.002) | -0.372 (2e-4) | **-0.581 (0.02)**† | -0.559 (0.02)† |
| DTLZ2 d15 m3 | -0.0941 (9e-4) | -0.100 (0.003) | -0.0991 (0.003) | -0.0993 (0.003) | -0.110 (0.004) | -0.129 (0.01)† | **-0.152 (0.01)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.373 (8e-4) | -0.373 (5e-4) | -0.373 (1e-3) | **-0.515 (0.03)**† | -0.466 (0.03)† |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0930 (1e-4) | -0.0935 (6e-4) | -0.0990 (0.004)† | -0.102 (0.003)† | -0.0975 (0.003)† | **-0.120 (0.01)**† |
| DTLZ2 m5 | 0.147 (0.003) | 0.120 (0.009)† | 0.132 (0.005) | **0.107 (0.005)**† | 0.110 (0.009)† | 0.140 (0.006) | 0.135 (0.008) |
| Penicillin | 5.68 (0.004) | 5.65 (0.02)† | 5.64 (0.01)† | 5.63 (0.02)† | 5.66 (0.008)† | **5.62 (0.01)**† | 5.63 (0.03)† |
| VehicleSafety | 1.21 (0.01) | 0.688 (0.08) | **0.345 (0.07)**† | 0.786 (0.07) | 0.570 (0.08)† | 0.401 (0.06)† | 0.452 (0.06)† |
| WFG6 m4 | 2.65 (0.003) | 2.60 (0.02) | 2.63 (0.01) | 2.64 (0.005) | 2.58 (0.02) | **2.24 (0.01)**† | 2.32 (0.03) |
| ZDT1 d20 m2 | 1.41 (0.02) | 1.28 (0.03)† | 1.25 (0.05)† | 1.36 (0.03) | 1.37 (0.02) | 1.18 (0.04)† | **1.15 (0.05)**† |
| ZDT2 d20 m2 | 1.62 (0.01) | 1.43 (0.06)† | **1.39 (0.03)**† | 1.54 (0.02) | 1.49 (0.04)† | 1.45 (0.04)† | 1.41 (0.05)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 2.15 (0.04) | 2.18 (0.03) | 2.13 (0.03) | 2.09 (0.03) | **1.94 (0.02)**† | 2.11 (0.03) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.53 (0.08) | 4.70 (0.04) | 4.66 (0.07) | 4.54 (0.07) | **4.31 (0.01)**† | 4.35 (0.02)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 7.05 (0.1) | 7.06 (0.09) | 7.17 (0.05) | 6.52 (0.09)† | **6.50 (0.07)**† | 6.54 (0.06)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 4.82 (0.05) | 4.92 (0.06) | 4.94 (0.05) | 4.73 (0.07)† | 4.69 (0.04)† | **4.66 (0.03)**† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 7.53 (0.06) | 7.60 (0.04) | 7.67 (0.02) | 7.11 (0.1)† | 6.96 (0.05)† | **6.93 (0.06)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.376 (0.004) | -0.373 (7e-4) | -0.374 (0.002) | -0.373 (8e-4) | **-0.515 (0.03)**† | -0.442 (0.03)† |
| DTLZ2 d15 m3 | -0.0941 (9e-4)† | -0.0934 (5e-4) | -0.0981 (0.002)† | -0.0958 (0.002)† | -0.0997 (0.003)† | -0.121 (0.01)† | **-0.132 (0.02)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | **-0.525 (0.03)**† | -0.437 (0.03) |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0931 (2e-4)† | -0.0939 (7e-4)† | -0.0950 (0.001)† | **-0.118 (0.01)**† | -0.107 (0.009)† |
| DTLZ2 m5 | 0.147 (0.003) | 0.138 (0.005)† | 0.139 (0.006)† | **0.124 (0.008)**† | 0.125 (0.006)† | 0.147 (0.004) | 0.135 (0.007)† |
| Penicillin | 5.68 (0.004) | 5.64 (0.03)† | **5.61 (0.02)**† | 5.62 (0.02)† | 5.65 (0.01)† | 5.62 (0.02)† | 5.63 (0.02)† |
| VehicleSafety | 1.21 (0.01) | 0.491 (0.06)† | 0.469 (0.09)† | 0.618 (0.06) | 0.524 (0.07)† | 0.414 (0.05)† | **0.397 (0.03)**† |
| WFG6 m4 | 2.65 (0.003) | 2.63 (0.01) | 2.59 (0.02) | 2.61 (0.01) | 2.42 (0.05) | **2.23 (0.02)**† | 2.28 (0.02)† |
| ZDT1 d20 m2 | 1.41 (0.02) | 1.19 (0.08)† | **0.969 (0.04)**† | 1.23 (0.04) | 1.16 (0.04) | 1.19 (0.04) | 1.24 (0.02) |
| ZDT2 d20 m2 | 1.62 (0.01) | 1.36 (0.04) | **1.06 (0.08)**† | 1.41 (0.04) | 1.26 (0.05)† | 1.46 (0.03) | 1.51 (0.01) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 2.28 (0.009) | 2.04 (0.01) | **1.99 (0.01)**† | 2.04 (0.02)† | 2.13 (0.02) | 2.01 (0.006) | 2.06 (0.02) |
| DTLZ1 d15 m2 | 5.09 (0.02) | 4.60 (0.06) | 4.79 (0.06) | 4.83 (0.05) | 5.08 (0.02) | 4.45 (0.01)† | **4.44 (0.02)**† |
| DTLZ1 d15 m3 | 7.52 (0.02) | 7.12 (0.08) | 7.26 (0.05) | 7.30 (0.04) | 7.42 (0.09) | 6.59 (0.02) | **6.52 (0.04)**† |
| DTLZ1 d20 m2 | 5.20 (0.005) | 4.96 (0.04) | 5.03 (0.03) | 4.95 (0.04) | 5.20 (0.002) | 4.70 (0.01)† | **4.66 (0.009)**† |
| DTLZ1 d20 m3 | 7.75 (0.008) | 7.41 (0.06) | 7.58 (0.04) | 7.60 (0.04) | 7.61 (0.07) | 7.03 (0.04) | **6.89 (0.01)**† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.378 (0.006) | -0.375 (0.001) | -0.390 (0.008) | -0.372 (0) | -0.574 (0.003)† | **-0.580 (0.004)**† |
| DTLZ2 d15 m3 | -0.0967 (0.002) | -0.119 (0.009) | -0.108 (0.007) | -0.137 (0.01) | -0.101 (0.003) | -0.175 (0.003) | **-0.197 (0.006)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.378 (0.004) | -0.372 (0) | -0.566 (0.004)† | **-0.569 (0.004)**† |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0983 (0.004) | -0.0948 (0.001) | -0.0990 (0.003) | -0.0932 (3e-4) | -0.142 (0.01) | **-0.188 (0.003)**† |
| DTLZ2 m5 | 0.140 (0.002) | 0.103 (0.01)† | 0.0989 (0.007)† | 0.121 (0.005) | **0.0961 (0.007)**† | 0.148 (0.005) | 0.102 (0.002)† |
| Penicillin | 5.66 (0.005) | 5.56 (0.02)† | 5.58 (0.02)† | 5.57 (0.01)† | 5.58 (0.02)† | **5.54 (0.01)**† | **5.54 (0.02)**† |
| VehicleSafety | 1.24 (0.01) | 0.867 (0.007) | **0.817 (0.006)**† | 0.846 (0.01) | 0.881 (0.03) | 0.886 (0.02) | 0.856 (0.008) |
| WFG6 m4 | 2.65 (0.003) | 2.62 (0.01) | 2.61 (0.01) | 2.63 (0.01) | 2.53 (0.02)† | **2.51 (0.01)**† | **2.51 (0.01)**† |
| ZDT1 d20 m2 | 1.40 (0.02) | 1.20 (0.05)† | **1.08 (0.05)**† | 1.38 (0.02) | 1.32 (0.04) | 1.15 (0.04)† | 1.19 (0.03)† |
| ZDT2 d20 m2 | 1.63 (0.01) | **1.36 (0.04)**† | 1.45 (0.04)† | 1.41 (0.03)† | 1.42 (0.05)† | 1.39 (0.04)† | 1.44 (0.03)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 2.28 (0.008) | 2.19 (0.02) | 2.15 (0.02) | 2.17 (0.02) | 2.23 (0.01) | **1.96 (0.02)**† | 2.07 (0.01) |
| DTLZ1 d15 m2 | 5.08 (0.02) | 4.89 (0.02) | 4.84 (0.05) | 4.94 (0.02) | 5.07 (0.01) | **4.26 (0.009)**† | 4.27 (0.01)† |
| DTLZ1 d15 m3 | 7.50 (0.02) | 7.29 (0.03) | 7.30 (0.03) | 7.40 (0.02) | 7.49 (0.02) | **6.21 (0.02)**† | 6.27 (0.02)† |
| DTLZ1 d20 m2 | 5.19 (0.006) | 5.13 (0.02) | 5.16 (0.009) | 5.13 (0.04) | 5.19 (0.004) | **4.56 (0.01)**† | 4.58 (0.008)† |
| DTLZ1 d20 m3 | 7.74 (0.009) | 7.65 (0.02) | 7.63 (0.02) | 7.70 (0.008) | 7.74 (0.009) | **6.70 (0.01)**† | 6.79 (0.04)† |
| DTLZ2 d15 m2 | -0.372 (0) | -0.372 (0) | -0.373 (7e-4) | -0.378 (0.004) | -0.374 (0.002) | -0.388 (0.02) | **-0.492 (0.03)**† |
| DTLZ2 d15 m3 | -0.0941 (9e-4)† | -0.0964 (0.002)† | -0.0939 (6e-4)† | -0.107 (0.006)† | -0.0952 (0.002)† | -0.102 (0.006)† | **-0.123 (0.02)**† |
| DTLZ2 d20 m2 | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.372 (0) | -0.389 (0.02) | **-0.524 (0.03)**† |
| DTLZ2 d20 m3 | -0.0929 (0) | -0.0929 (0) | -0.0929 (0) | -0.100 (0.003)† | -0.0929 (0) | -0.0929 (0) | **-0.115 (0.01)**† |
| DTLZ2 m5 | 0.147 (0.003) | 0.145 (0.004) | 0.140 (0.007) | **0.109 (0.01)**† | 0.125 (0.006)† | 0.153 (0.002) | 0.154 (0.002) |
| Penicillin | 5.68 (0.004) | 5.63 (0.01)† | 5.64 (0.006)† | 5.66 (0.02)† | 5.66 (0.01) | **5.60 (0.02)**† | 5.64 (0.02)† |
| VehicleSafety | 1.21 (0.01) | 0.849 (0.09) | 0.509 (0.08)† | 0.597 (0.07) | 0.947 (0.07) | **0.375 (0.03)**† | 0.694 (0.08) |
| WFG6 m4 | 2.65 (0.003) | 2.66 (0.007) | 2.65 (0.009) | 2.64 (0.01) | 2.63 (0.009) | **2.25 (0.01)**† | 2.26 (0.02)† |
| ZDT1 d20 m2 | 1.41 (0.02) | 1.31 (0.06) | 1.32 (0.04) | 1.40 (0.03) | 1.41 (0.02) | **1.12 (0.06)**† | 1.13 (0.08)† |
| ZDT2 d20 m2 | 1.62 (0.01) | 1.53 (0.03) | 1.48 (0.04)† | 1.57 (0.03) | 1.62 (0.01) | **1.19 (0.2)**† | 1.46 (0.05)† |

Table 33: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.05$, oracle values (11), which score every evaluated design: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0192 (8e-4)† | **0.0167 (0.001)**† | 0.0497 (0.006) | 0.0632 (0.001) | 0.0375 (0.002) | 0.0597 (0.003) |
| DTLZ1 d15 m2 | 646 (30) | 253 (5) | 248 (3) | **238 (3)**† | 261 (10) | 242 (3)† | 248 (6)† |
| DTLZ1 d15 m3 | 512 (30) | 223 (6)† | 220 (5)† | **207 (6)**† | 263 (10) | 221 (6)† | 221 (4)† |
| DTLZ1 d20 m2 | 939 (30) | 337 (4)† | 335 (6)† | **328 (2)**† | 388 (20) | 360 (10) | 344 (4) |
| DTLZ1 d20 m3 | 775 (20) | 288 (8)† | **287 (7)**† | 296 (10)† | 357 (10) | 301 (8)† | 306 (9)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.190 (0.01)† | 0.215 (0.02)† | **0.184 (0.01)**† | 0.741 (0.02) | 0.308 (0.004) | 0.628 (0.07) |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.522 (0.07)† | 0.467 (0.06)† | 0.521 (0.05) | 0.752 (0.02) | **0.393 (0.008)**† | 0.760 (0.04) |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.338 (0.03)† | 0.371 (0.07)† | 0.323 (0.02)† | 1.03 (0.02) | **0.320 (0.002)**† | 1.08 (0.02) |
| DTLZ2 d20 m3 | 0.982 (0.02)† | **0.645 (0.1)**† | 0.780 (0.09)† | 0.683 (0.05)† | 1.01 (0.03) | 0.743 (0.1)† | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.482 (0.05)† | **0.463 (0.04)**† | 0.564 (0.03) | 0.696 (0.02) | 0.526 (0.01) | 0.472 (0.003) |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 0.00226 (7e-4) | 1.28e-4 (1e-4)† | 3.24e-5 (2e-5) |
| Penicillin | 0.0698 (0.001) | 0.0555 (0.005)† | 0.0649 (0.004)† | **0.0535 (0.002)**† | 0.0650 (0.002) | 0.0822 (0.004) | 0.0853 (0.002) |
| VehicleSafety | 0.181 (0.009) | 0.00290 (4e-4) | **5.36e-4 (1e-4)**† | 0.0123 (0.001) | 0.00881 (0.001) | 0.0101 (0.002) | 0.0102 (8e-4) |
| WFG6 m4 | 0.123 (0.002) | **0.0365 (0.002)**† | 0.0408 (0.003)† | 0.0866 (0.003) | 0.0716 (0.005) | 0.0557 (0.002) | 0.0461 (0.003) |
| ZDT1 d20 m2 | 2.01 (0.08) | **0.0906 (0.02)**† | 0.120 (0.02)† | 0.129 (0.03)† | 0.359 (0.02) | 0.446 (0.06) | 0.389 (0.08) |
| ZDT2 d20 m2 | 3.23 (0.08) | **0.0632 (0.02)**† | 0.0842 (0.01)† | 0.111 (0.03)† | 0.743 (0.06) | 0.749 (0.2) | 0.595 (0.1) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0193 (8e-4)† | **0.0183 (0.002)**† | 0.0481 (0.005) | 0.0668 (0.001) | 0.0423 (0.004) | 0.0552 (0.003) |
| DTLZ1 d15 m2 | 646 (30) | 252 (3) | 254 (7) | **236 (3)**† | 261 (7) | 245 (2) | 256 (10)† |
| DTLZ1 d15 m3 | 512 (30) | 215 (5)† | 230 (6) | **204 (6)**† | 247 (10) | 215 (6)† | 223 (5) |
| DTLZ1 d20 m2 | 939 (30) | 333 (3)† | **329 (3)**† | 336 (5)† | 343 (7)† | 341 (4) | 364 (10) |
| DTLZ1 d20 m3 | 775 (20) | 290 (6)† | **278 (8)**† | 295 (10)† | 356 (10) | 322 (6) | 325 (10) |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.260 (0.03)† | 0.241 (0.03)† | **0.232 (0.03)**† | 0.731 (0.02) | 0.304 (0.005)† | 0.758 (0.02) |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.471 (0.07)† | 0.469 (0.06)† | 0.442 (0.04)† | 0.756 (0.02) | **0.397 (0.004)**† | 0.790 (0.02) |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.339 (0.03)† | 0.313 (0.04)† | **0.287 (0.04)**† | 1.03 (0.03) | 0.317 (0.002) | 1.02 (0.08) |
| DTLZ2 d20 m3 | 0.982 (0.02)† | 0.766 (0.1)† | 0.781 (0.1)† | 0.684 (0.06)† | 1.05 (0.02) | **0.633 (0.1)**† | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.484 (0.05)† | 0.484 (0.05)† | 0.543 (0.03)† | 0.695 (0.02) | 0.568 (0.04) | **0.463 (0.006)**† |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 9.92e-4 (4e-4) | 3.02e-5 (2e-5) | 5.69e-5 (3e-5) |
| Penicillin | 0.0698 (0.001) | 0.0659 (0.002) | 0.0646 (0.003) | **0.0534 (0.002)**† | 0.0612 (0.004) | 0.0777 (0.004) | 0.0875 (0.004) |
| VehicleSafety | 0.181 (0.009) | 0.00303 (1e-3) | **6.50e-4 (1e-4)**† | 0.0114 (8e-4) | 0.00860 (7e-4) | 0.0173 (0.005) | 0.0132 (0.001) |
| WFG6 m4 | 0.123 (0.002) | **0.0335 (0.002)**† | 0.0345 (0.002)† | 0.0907 (0.004) | 0.0711 (0.006) | 0.0546 (0.002) | 0.0434 (0.002) |
| ZDT1 d20 m2 | 2.01 (0.08) | **0.0819 (0.01)**† | 0.0854 (0.009)† | 0.0946 (0.02)† | 0.377 (0.02) | 0.413 (0.1) | 0.519 (0.08) |
| ZDT2 d20 m2 | 3.23 (0.08) | **0.0397 (0.006)**† | 0.0863 (0.01) | 0.228 (0.03) | 0.803 (0.06) | 0.564 (0.09) | 0.828 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.107 (0.002) | 0.0357 (0.002) | **0.0287 (0.001)**† | 0.0575 (0.004) | 0.0769 (0.001) | 0.0511 (0.002) | 0.0588 (0.002) |
| DTLZ1 d15 m2 | 658 (40) | 325 (10)† | 302 (2)† | **297 (3)**† | 307 (5)† | 300 (1)† | 309 (6)† |
| DTLZ1 d15 m3 | 523 (30) | 249 (6)† | 252 (5)† | **240 (4)**† | 278 (10) | 259 (7)† | 274 (7) |
| DTLZ1 d20 m2 | 993 (30) | 398 (3) | 393 (2)† | **382 (4)**† | 411 (10) | 403 (8) | 414 (10) |
| DTLZ1 d20 m3 | 829 (30) | 322 (5)† | 327 (7)† | **318 (9)**† | 389 (20) | 331 (7)† | 380 (4) |
| DTLZ2 d15 m2 | 0.601 (0.02) | 0.242 (0.03)† | 0.217 (0.01)† | **0.205 (0.02)**† | 0.659 (0.03) | 0.305 (0.003) | 0.670 (0.05) |
| DTLZ2 d15 m3 | 0.626 (0.02) | 0.423 (0.05)† | 0.432 (0.05)† | 0.439 (0.03)† | 0.641 (0.03) | **0.384 (0.005)**† | 0.571 (0.06) |
| DTLZ2 d20 m2 | 0.920 (0.02) | 0.589 (0.06) | 0.410 (0.06)† | 0.363 (0.03)† | 0.986 (0.02) | **0.326 (0.004)**† | 1.02 (0.02) |
| DTLZ2 d20 m3 | 0.944 (0.02) | 0.701 (0.1)† | 0.837 (0.08) | 0.807 (0.05) | 0.964 (0.04) | **0.411 (0.003)**† | 1.06 (0.01) |
| DTLZ2 m5 | 0.607 (0.008) | 0.527 (0.04)† | 0.561 (0.04)† | 0.564 (0.03) | 0.615 (0.02) | 0.506 (0.009) | **0.445 (0.006)**† |
| DTLZ7 m5 | 0.206 (0.008) | 5.47e-6 (2e-6) | **0 (0)**† | 5.15e-8 (5e-8)† | 0.0180 (0.005) | 3.98e-8 (4e-8)† | 5.72e-5 (5e-5) |
| Penicillin | 0.0730 (0.001) | 0.0542 (0.004)† | 0.0675 (0.002) | **0.0521 (0.002)**† | 0.0676 (0.002) | 0.0852 (0.003) | 0.0811 (0.002) |
| VehicleSafety | 0.189 (0.01) | 0.0194 (4e-4) | **0.0167 (6e-5)**† | 0.0285 (1e-3) | 0.0236 (9e-4) | 0.0302 (0.001) | 0.0298 (0.001) |
| WFG6 m4 | 0.122 (0.002) | 0.0780 (0.002)† | **0.0762 (0.002)**† | 0.105 (0.002) | 0.0981 (0.002) | 0.0897 (8e-4) | 0.0918 (0.002) |
| ZDT1 d20 m2 | 1.98 (0.08) | 0.421 (0.04) | **0.189 (0.02)**† | 0.236 (0.03)† | 0.443 (0.02) | 0.354 (0.04) | 0.420 (0.04) |
| ZDT2 d20 m2 | 3.23 (0.08) | 0.595 (0.02) | **0.279 (0.008)**† | 0.380 (0.05)† | 0.854 (0.07) | 0.478 (0.08) | 0.924 (0.2) |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0297 (0.002) | **0.0197 (0.002)**† | 0.0569 (0.004) | 0.0616 (0.002) | 0.0312 (0.001) | 0.0513 (0.002) |
| DTLZ1 d15 m2 | 646 (30) | 258 (9)† | 249 (8)† | 244 (2)† | 270 (8) | 243 (2)† | **242 (3)**† |
| DTLZ1 d15 m3 | 512 (30) | 217 (7)† | 208 (5)† | 217 (9)† | 260 (10) | **201 (2)**† | 214 (6)† |
| DTLZ1 d20 m2 | 939 (30) | 339 (5) | 337 (4) | **327 (2)**† | 369 (20) | 338 (7)† | 342 (5) |
| DTLZ1 d20 m3 | 775 (20) | 296 (7)† | **290 (5)**† | 294 (9)† | 374 (20) | 300 (9)† | 310 (10) |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.187 (0.03)† | **0.158 (0.008)**† | 0.212 (0.02) | 0.712 (0.02) | 0.305 (0.002) | 0.758 (0.02) |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.457 (0.07)† | **0.319 (0.02)**† | 0.426 (0.05)† | 0.678 (0.03) | 0.388 (0.003)† | 0.639 (0.07) |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.387 (0.1)† | 0.505 (0.1)† | 0.375 (0.05)† | 0.989 (0.04) | **0.314 (0.002)**† | 1.08 (0.02) |
| DTLZ2 d20 m3 | 0.982 (0.02) | 0.752 (0.1) | 0.586 (0.1)† | 0.667 (0.07) | 0.979 (0.03) | **0.399 (0.003)**† | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.582 (0.05)† | 0.533 (0.05)† | 0.606 (0.03) | 0.640 (0.03) | 0.597 (0.03) | **0.453 (0.009)**† |
| DTLZ7 m5 | 0.199 (0.009) | **0 (0)**† | **0 (0)**† | **0 (0)**† | 0.00285 (6e-4) | 2.34e-5 (2e-5)† | 3.79e-6 (4e-6)† |
| Penicillin | 0.0698 (0.001) | 0.0631 (0.004)† | 0.0661 (0.002) | **0.0548 (0.002)**† | 0.0633 (0.003)† | 0.0893 (0.004) | 0.0896 (0.004) |
| VehicleSafety | 0.181 (0.009) | 0.00286 (2e-4) | **5.77e-4 (9e-5)**† | 0.0115 (0.002) | 0.00961 (0.001) | 0.0126 (0.001) | 0.00823 (1e-3) |
| WFG6 m4 | 0.123 (0.002) | 0.0396 (0.001)† | **0.0362 (0.002)**† | 0.0880 (0.004) | 0.0762 (0.003) | 0.0540 (0.002) | 0.0409 (0.001)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 0.142 (0.01) | **0.0777 (0.007)**† | 0.217 (0.03) | 0.331 (0.01) | 0.229 (0.05) | 0.336 (0.05) |
| ZDT2 d20 m2 | 3.23 (0.08) | 0.0828 (0.02)† | **0.0563 (0.009)**† | 0.167 (0.04)† | 0.969 (0.07) | 0.349 (0.04) | 0.741 (0.1) |

Table 34: Terminal $\mathrm{IGD}^{+}$ at $\eta = 0.2$, oracle values (11), which score every evaluated design: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0773 (0.004) | **0.0596 (0.005)**† | 0.0896 (0.006) | 0.0790 (0.004) | 0.0673 (0.006)† | 0.0821 (0.007) |
| DTLZ1 d15 m2 | 646 (30) | 287 (30)† | **261 (6)**† | 286 (20)† | 366 (50) | **261 (4)**† | 270 (8)† |
| DTLZ1 d15 m3 | 512 (30) | 270 (20) | 257 (20)† | 321 (20) | 240 (10)† | 242 (10)† | **229 (5)**† |
| DTLZ1 d20 m2 | 939 (30) | **364 (10)**† | 393 (20)† | 392 (20)† | 430 (20) | 371 (10)† | 378 (10)† |
| DTLZ1 d20 m3 | 775 (20) | 440 (40) | 469 (40) | 550 (30) | 377 (30)† | **333 (10)**† | 343 (10)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.650 (0.03) | 0.596 (0.05) | 0.447 (0.02) | 0.464 (0.02) | **0.312 (0.003)**† | 0.507 (0.08)† |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.720 (0.03) | 0.692 (0.03) | 0.569 (0.04)† | **0.521 (0.02)**† | 0.547 (0.06)† | 0.564 (0.06)† |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.855 (0.05) | 0.782 (0.06) | 0.621 (0.06) | 0.619 (0.03) | **0.319 (0.003)**† | 0.803 (0.1) |
| DTLZ2 d20 m3 | 0.982 (0.02) | 0.985 (0.03) | 0.979 (0.03) | 0.724 (0.03)† | **0.689 (0.03)**† | 0.814 (0.1)† | 0.825 (0.1)† |
| DTLZ2 m5 | 0.643 (0.01) | 0.646 (0.03) | 0.641 (0.02) | 0.545 (0.02)† | **0.508 (0.02)**† | 0.618 (0.04) | 0.556 (0.04)† |
| DTLZ7 m5 | 0.199 (0.009) | 0.00634 (0.004) | 0.00180 (9e-4)† | 0.0270 (0.009) | 0.0172 (0.004) | **7.87e-4 (5e-4)**† | 0.00225 (7e-4)† |
| Penicillin | 0.0698 (0.001) | 0.0676 (0.003)† | 0.0659 (0.003)† | **0.0629 (0.003)**† | 0.0633 (0.003)† | 0.0833 (0.004) | 0.0832 (0.003) |
| VehicleSafety | 0.181 (0.009) | 0.0139 (0.002) | **0.00545 (0.001)**† | 0.0176 (0.003) | 0.0110 (0.001) | 0.0136 (0.001) | 0.0181 (0.004) |
| WFG6 m4 | 0.123 (0.002) | 0.104 (0.006) | 0.108 (0.005) | 0.113 (0.002) | 0.0863 (0.005) | **0.0576 (0.002)**† | 0.0632 (0.003)† |
| ZDT1 d20 m2 | 2.01 (0.08) | **0.688 (0.05)**† | 0.700 (0.1)† | 1.31 (0.1) | 0.991 (0.2)† | 1.01 (0.1) | 1.04 (0.1) |
| ZDT2 d20 m2 | 3.23 (0.08) | 0.797 (0.2)† | **0.717 (0.2)**† | 1.43 (0.2) | 0.967 (0.1)† | 1.75 (0.2) | 1.54 (0.3) |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0729 (0.005)† | **0.0626 (0.005)**† | 0.0766 (0.005)† | 0.0724 (0.005)† | **0.0626 (0.005)**† | 0.0817 (0.005) |
| DTLZ1 d15 m2 | 646 (30) | 272 (10)† | 306 (10) | 270 (10)† | 305 (30)† | 276 (10)† | **265 (7)**† |
| DTLZ1 d15 m3 | 512 (30) | 251 (20)† | 235 (10)† | 309 (20) | **227 (6)**† | 232 (7)† | 240 (6)† |
| DTLZ1 d20 m2 | 939 (30) | 385 (20)† | 392 (20)† | 409 (20)† | 429 (70)† | 365 (9)† | **363 (10)**† |
| DTLZ1 d20 m3 | 775 (20) | 426 (30) | 414 (40)† | 527 (40) | 383 (20)† | **333 (7)**† | 360 (10)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.634 (0.04) | 0.547 (0.04) | 0.491 (0.05)† | 0.576 (0.04) | **0.344 (0.04)**† | 0.626 (0.07) |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.714 (0.03) | 0.696 (0.02) | 0.608 (0.04)† | 0.526 (0.03)† | **0.490 (0.05)**† | 0.654 (0.06)† |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.954 (0.05) | 0.695 (0.06) | 0.610 (0.07) | 0.714 (0.05) | **0.409 (0.08)**† | 0.932 (0.1) |
| DTLZ2 d20 m3 | 0.982 (0.02) | 0.987 (0.03) | 1.00 (0.02) | 0.753 (0.05)† | **0.689 (0.03)**† | 0.938 (0.09) | 1.07 (0.02) |
| DTLZ2 m5 | 0.643 (0.01) | 0.663 (0.02) | 0.641 (0.02) | 0.547 (0.02)† | **0.516 (0.02)**† | 0.638 (0.03) | 0.570 (0.04)† |
| DTLZ7 m5 | 0.199 (0.009) | 0.00261 (9e-4)† | **1.76e-4 (7e-5)**† | 0.0105 (0.006) | 0.0124 (0.003) | 3.86e-4 (2e-4)† | 0.00368 (0.002)† |
| Penicillin | 0.0698 (0.001) | 0.0626 (0.004)† | 0.0670 (0.003) | **0.0560 (0.003)**† | 0.0665 (0.002) | 0.0826 (0.005) | 0.0905 (0.005) |
| VehicleSafety | 0.181 (0.009) | 0.0126 (0.001) | **0.00392 (8e-4)**† | 0.0106 (0.002) | 0.0129 (0.003) | 0.0140 (0.001) | 0.0141 (0.001) |
| WFG6 m4 | 0.123 (0.002) | 0.104 (0.006) | 0.0901 (0.008) | 0.112 (0.002) | 0.0842 (0.006) | **0.0569 (0.002)**† | 0.0648 (0.004)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 0.755 (0.2)† | **0.382 (0.04)**† | 0.803 (0.1) | 0.670 (0.1) | 0.837 (0.1) | 1.28 (0.1) |
| ZDT2 d20 m2 | 3.23 (0.08) | 0.509 (0.1)† | **0.355 (0.08)**† | 1.07 (0.2) | 0.860 (0.09) | 1.56 (0.2) | 1.93 (0.2) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.107 (0.002) | 0.0841 (0.005) | **0.0587 (0.005)**† | 0.0727 (0.007)† | 0.0771 (0.002) | 0.0747 (0.003) | 0.0835 (0.004) |
| DTLZ1 d15 m2 | 658 (40) | 370 (20) | 355 (20) | 318 (10)† | 453 (50) | **296 (3)**† | 304 (5)† |
| DTLZ1 d15 m3 | 523 (30) | 306 (10) | 305 (20) | 336 (20) | 375 (30) | **263 (5)**† | 280 (10)† |
| DTLZ1 d20 m2 | 993 (30) | 438 (20)† | 440 (20)† | 443 (20)† | 908 (50) | **416 (10)**† | 418 (9)† |
| DTLZ1 d20 m3 | 829 (30) | 492 (30) | 481 (30) | 520 (40) | 579 (50) | 358 (9)† | **355 (10)**† |
| DTLZ2 d15 m2 | 0.601 (0.02) | 0.493 (0.03) | 0.452 (0.04) | 0.437 (0.03) | 0.457 (0.03) | **0.308 (0.006)**† | 0.460 (0.07) |
| DTLZ2 d15 m3 | 0.626 (0.02) | 0.604 (0.02) | 0.601 (0.03) | 0.496 (0.03) | 0.458 (0.02) | **0.390 (0.005)**† | 0.391 (0.007)† |
| DTLZ2 d20 m2 | 0.920 (0.02) | 0.823 (0.05) | 0.843 (0.05) | 0.663 (0.03) | 0.632 (0.03) | **0.322 (0.005)**† | 0.800 (0.1) |
| DTLZ2 d20 m3 | 0.944 (0.02) | 0.934 (0.03) | 0.893 (0.04) | 0.679 (0.03) | 0.570 (0.02) | **0.418 (0.005)**† | 0.473 (0.07)† |
| DTLZ2 m5 | 0.607 (0.008) | 0.586 (0.02) | 0.581 (0.01) | 0.469 (0.02)† | 0.468 (0.01)† | 0.590 (0.02) | **0.460 (0.01)**† |
| DTLZ7 m5 | 0.206 (0.008) | 9.90e-5 (5e-5)† | **1.52e-5 (1e-5)**† | 4.77e-5 (3e-5)† | 0.0249 (0.005) | 1.70e-4 (9e-5)† | 4.30e-4 (2e-4) |
| Penicillin | 0.0730 (0.001) | 0.0594 (0.003)† | 0.0686 (0.002) | **0.0591 (0.002)**† | 0.0678 (0.002) | 0.0810 (0.002) | 0.0828 (0.002) |
| VehicleSafety | 0.189 (0.01) | 0.0299 (0.002) | **0.0192 (5e-4)**† | 0.0273 (0.001) | 0.0260 (0.002) | 0.0363 (0.005) | 0.0308 (0.002) |
| WFG6 m4 | 0.122 (0.002) | 0.107 (0.005)† | 0.108 (0.003) | 0.107 (0.005)† | 0.0999 (0.004)† | **0.0970 (0.003)**† | 0.104 (0.002)† |
| ZDT1 d20 m2 | 1.98 (0.08) | 0.796 (0.07) | 0.627 (0.05)† | 1.02 (0.1) | 0.864 (0.09) | **0.613 (0.06)**† | 0.798 (0.1)† |
| ZDT2 d20 m2 | 3.23 (0.08) | 1.15 (0.1)† | 1.18 (0.3)† | 1.05 (0.1)† | **1.02 (0.08)**† | 1.03 (0.2)† | 1.34 (0.2)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0784 (0.005) | 0.0739 (0.006) | 0.0727 (0.007)† | 0.0780 (0.003) | **0.0551 (0.006)**† | 0.0850 (0.005) |
| DTLZ1 d15 m2 | 646 (30) | 325 (30) | 319 (20)† | 289 (20)† | 580 (40) | **244 (3)**† | 251 (5)† |
| DTLZ1 d15 m3 | 512 (30) | 310 (20) | 302 (20) | 382 (30) | 488 (30) | **206 (4)**† | 216 (4)† |
| DTLZ1 d20 m2 | 939 (30) | 469 (30) | 489 (30) | 471 (40) | 880 (30) | **335 (2)**† | 342 (5)† |
| DTLZ1 d20 m3 | 775 (20) | 445 (30) | 490 (30) | 580 (40) | 745 (20) | **300 (8)**† | 305 (7)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.617 (0.02) | 0.625 (0.04) | 0.559 (0.04) | 0.513 (0.02) | **0.317 (0.002)**† | 0.465 (0.07)† |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.623 (0.04) | 0.654 (0.04)† | 0.589 (0.04)† | 0.484 (0.01) | 0.659 (0.07)† | **0.468 (0.06)**† |
| DTLZ2 d20 m2 | 0.980 (0.02)† | 0.938 (0.04)† | 0.960 (0.03)† | 0.804 (0.04)† | 0.755 (0.03)† | **0.567 (0.1)**† | 0.954 (0.08) |
| DTLZ2 d20 m3 | 0.982 (0.02) | 1.00 (0.02) | 0.982 (0.03) | 0.775 (0.03) | **0.628 (0.02)**† | 0.939 (0.08) | 0.755 (0.1)† |
| DTLZ2 m5 | 0.643 (0.01)† | 0.638 (0.02) | 0.644 (0.02) | 0.558 (0.03)† | 0.546 (0.02)† | 0.733 (0.02) | **0.538 (0.03)**† |
| DTLZ7 m5 | 0.199 (0.009) | 8.57e-5 (4e-5) | 2.66e-4 (2e-4) | 8.43e-5 (8e-5) | 0.0168 (0.004) | **0 (0)**† | 6.38e-4 (4e-4) |
| Penicillin | 0.0698 (0.001) | 0.0600 (0.006)† | 0.0635 (0.002)† | **0.0580 (0.003)**† | 0.0600 (0.002)† | 0.0877 (0.003) | 0.0837 (0.003) |
| VehicleSafety | 0.181 (0.009) | 0.0169 (0.005)† | **0.00806 (0.002)**† | 0.0188 (0.006) | 0.0132 (0.002)† | 0.0149 (0.001) | 0.0135 (0.002) |
| WFG6 m4 | 0.123 (0.002) | 0.122 (0.004) | 0.122 (0.003) | 0.109 (0.002) | 0.120 (0.005) | **0.0564 (0.003)**† | 0.0568 (0.003)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 0.915 (0.2) | 0.888 (0.1) | 1.31 (0.2) | 2.05 (0.08) | **0.430 (0.07)**† | 0.752 (0.1) |
| ZDT2 d20 m2 | 3.23 (0.08) | 0.671 (0.1)† | **0.572 (0.1)**† | 0.641 (0.1)† | 2.07 (0.2) | 1.04 (0.2) | 1.12 (0.2) |

Table 35: Terminal IGD$^+$ at $\eta = 0.4$, oracle values (11), which score every evaluated design: mean over ten seeds (standard error). The best mean of each row is in bold; † marks it and every method not significantly different from it (Mann–Whitney $U$, 5%); lower is better.

| problem | Sobol | qLogEHVI | qLogNEHVI | qLogNParEGO | qLB-JES | LF-NPHVI | LF-NEHVI |
|---|---|---|---|---|---|---|---|
| *constant noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.100 (0.007) | 0.0983 (0.007) | 0.104 (0.005) | 0.0995 (0.004) | **0.0777 (0.005)**† | 0.107 (0.006) |
| DTLZ1 d15 m2 | 646 (30) | 290 (20)† | 313 (20) | 397 (20) | 461 (50) | **256 (8)**† | 266 (7)† |
| DTLZ1 d15 m3 | 512 (30) | 326 (30) | 372 (30) | 413 (30) | 300 (40) | **228 (7)**† | 235 (9)† |
| DTLZ1 d20 m2 | 939 (30) | 540 (50) | 651 (60) | 624 (50) | 738 (70) | 381 (10)† | **361 (10)**† |
| DTLZ1 d20 m3 | 775 (20) | 586 (50) | 666 (30) | 633 (40) | 643 (50) | **356 (10)**† | 359 (10)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.572 (0.03) | 0.543 (0.03) | 0.537 (0.02) | 0.407 (0.02) | **0.322 (0.01)**† | 0.329 (0.01)† |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.597 (0.03) | 0.585 (0.03) | 0.567 (0.03)† | **0.501 (0.02)**† | 0.618 (0.06)† | 0.520 (0.05)† |
| DTLZ2 d20 m2 | 0.980 (0.02)† | 0.751 (0.05)† | 0.683 (0.06)† | 0.728 (0.04)† | 0.590 (0.03)† | **0.543 (0.1)**† | 0.553 (0.09)† |
| DTLZ2 d20 m3 | 0.982 (0.02) | 0.749 (0.04) | 0.754 (0.05) | 0.668 (0.05)† | **0.575 (0.02)**† | 0.986 (0.06) | 0.808 (0.1)† |
| DTLZ2 m5 | 0.643 (0.01) | 0.556 (0.02)† | 0.594 (0.02) | **0.519 (0.01)**† | 0.526 (0.03)† | 0.645 (0.03) | 0.631 (0.04)† |
| DTLZ7 m5 | 0.199 (0.009) | 0.0464 (0.01) | 0.0648 (0.02) | 0.0933 (0.02) | 0.0406 (0.01) | **0.00255 (1e-3)**† | 0.0305 (0.01)† |
| Penicillin | 0.0698 (0.001)† | 0.0672 (0.003)† | **0.0668 (0.004)**† | 0.0692 (0.004)† | **0.0668 (0.003)**† | 0.0875 (0.004) | 0.0845 (0.003) |
| VehicleSafety | 0.181 (0.009) | 0.0332 (0.009) | **0.0118 (0.001)**† | 0.0586 (0.02) | 0.0307 (0.01) | 0.0132 (0.002)† | 0.0199 (0.004) |
| WFG6 m4 | 0.123 (0.002) | 0.114 (0.004) | 0.122 (0.002) | 0.118 (0.003) | 0.107 (0.004) | **0.0616 (0.002)**† | 0.0737 (0.005)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 1.51 (0.1)† | 1.42 (0.2)† | 1.82 (0.1) | 1.85 (0.1) | 1.19 (0.1)† | **1.13 (0.1)**† |
| ZDT2 d20 m2 | 3.23 (0.08) | 2.08 (0.2)† | **1.72 (0.2)**† | 2.59 (0.2) | 2.35 (0.2)† | 2.03 (0.2)† | 1.87 (0.2)† |
| *contaminated noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0935 (0.006) | 0.0960 (0.005) | 0.0870 (0.007)† | 0.0851 (0.005)† | **0.0713 (0.006)**† | 0.0910 (0.009)† |
| DTLZ1 d15 m2 | 646 (30) | 319 (40)† | 342 (30) | 353 (30) | 308 (30)† | **261 (7)**† | 275 (9)† |
| DTLZ1 d15 m3 | 512 (30) | 332 (20) | 308 (20) | 362 (30) | 242 (20)† | **230 (3)**† | 231 (6)† |
| DTLZ1 d20 m2 | 939 (30) | 463 (40)† | 500 (40)† | 511 (50) | 445 (60)† | 400 (10)† | **395 (20)**† |
| DTLZ1 d20 m3 | 775 (20) | 542 (60)† | 554 (40) | 660 (30) | 397 (30)† | **351 (10)**† | 356 (10)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.595 (0.03)† | 0.645 (0.04)† | 0.550 (0.04)† | 0.539 (0.02)† | **0.455 (0.07)**† | 0.593 (0.07)† |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.712 (0.02) | 0.613 (0.03)† | 0.653 (0.03) | **0.551 (0.02)**† | 0.673 (0.06)† | 0.594 (0.06)† |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.877 (0.03) | 0.836 (0.05) | 0.842 (0.04) | 0.601 (0.06) | **0.443 (0.07)**† | 0.762 (0.1)† |
| DTLZ2 d20 m3 | 0.982 (0.02) | 0.889 (0.03) | 0.776 (0.05)† | 0.775 (0.04)† | **0.664 (0.03)**† | 0.825 (0.1)† | 0.926 (0.09)† |
| DTLZ2 m5 | 0.643 (0.01) | 0.607 (0.02)† | 0.619 (0.03)† | 0.564 (0.02)† | **0.557 (0.02)**† | 0.683 (0.03) | 0.633 (0.03)† |
| DTLZ7 m5 | 0.199 (0.009) | 0.0442 (0.01) | 0.0734 (0.02) | 0.0740 (0.02) | 0.0371 (0.01) | **5.89e-4 (4e-4)**† | 0.0179 (0.009) |
| Penicillin | 0.0698 (0.001) | 0.0652 (0.004)† | 0.0649 (0.003)† | **0.0618 (0.003)**† | 0.0685 (0.003)† | 0.0853 (0.002) | 0.0906 (0.004) |
| VehicleSafety | 0.181 (0.009) | 0.0176 (0.003)† | 0.0233 (0.009)† | 0.0284 (0.007)† | 0.0261 (0.007)† | 0.0159 (0.002)† | **0.0153 (0.002)**† |
| WFG6 m4 | 0.123 (0.002) | 0.125 (0.003) | 0.117 (0.005) | 0.119 (0.004) | 0.0882 (0.007) | **0.0583 (0.002)**† | 0.0637 (0.003)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 1.38 (0.2) | **0.725 (0.06)**† | 1.35 (0.1) | 1.13 (0.1) | 1.24 (0.1) | 1.33 (0.05) |
| ZDT2 d20 m2 | 3.23 (0.08) | 1.58 (0.2) | **0.835 (0.1)**† | 1.82 (0.2) | 1.22 (0.2)† | 2.07 (0.2) | 2.33 (0.09) |
| *hidden inputs noise* | | | | | | | |
| CarSideImpact | 0.107 (0.002) | 0.101 (0.004) | **0.0748 (0.004)**† | 0.102 (0.007) | 0.0938 (0.003) | 0.0874 (0.002) | 0.0976 (0.006) |
| DTLZ1 d15 m2 | 658 (40) | 351 (20) | 411 (30) | 438 (20) | 649 (30) | 307 (8)† | **306 (5)**† |
| DTLZ1 d15 m3 | 523 (30) | 347 (40) | 376 (30) | 388 (20) | 504 (40) | 275 (6)† | **260 (7)**† |
| DTLZ1 d20 m2 | 993 (30) | 512 (30) | 575 (30) | 526 (40) | 993 (30) | **396 (7)**† | 397 (6)† |
| DTLZ1 d20 m3 | 829 (30) | 473 (40) | 589 (40) | 621 (40) | 648 (60) | 372 (10)† | **339 (9)**† |
| DTLZ2 d15 m2 | 0.601 (0.02) | 0.489 (0.03) | 0.472 (0.03) | 0.398 (0.03) | 0.523 (0.02) | 0.325 (0.003)† | **0.322 (0.005)**† |
| DTLZ2 d15 m3 | 0.626 (0.02) | 0.496 (0.03) | 0.501 (0.02) | 0.428 (0.02)† | 0.523 (0.02) | 0.420 (0.006) | **0.385 (0.008)**† |
| DTLZ2 d20 m2 | 0.920 (0.02) | 0.672 (0.03) | 0.641 (0.03) | 0.527 (0.04) | 0.705 (0.04) | 0.335 (0.004)† | **0.331 (0.004)**† |
| DTLZ2 d20 m3 | 0.944 (0.02) | 0.726 (0.04) | 0.704 (0.03) | 0.600 (0.03) | 0.672 (0.02) | 0.594 (0.08) | **0.402 (0.004)**† |
| DTLZ2 m5 | 0.607 (0.008) | 0.497 (0.03)† | 0.485 (0.01)† | 0.530 (0.01) | 0.471 (0.02)† | 0.673 (0.03) | **0.456 (0.006)**† |
| DTLZ7 m5 | 0.206 (0.008) | 0.00108 (7e-4)† | 1.26e-4 (6e-5)† | 1.76e-4 (9e-5)† | 0.0469 (0.009) | **7.34e-5 (3e-5)**† | 0.00538 (0.004) |
| Penicillin | 0.0730 (0.001)† | 0.0697 (0.004)† | **0.0673 (0.003)**† | 0.0694 (0.003)† | 0.0701 (0.002)† | 0.0814 (0.002) | 0.0755 (0.002) |
| VehicleSafety | 0.189 (0.01) | 0.0348 (0.002) | **0.0255 (0.001)**† | 0.0298 (0.001)† | 0.0434 (0.006) | 0.0385 (0.005) | 0.0331 (0.001) |
| WFG6 m4 | 0.122 (0.002) | 0.124 (0.004) | 0.122 (0.004) | 0.121 (0.005) | **0.102 (0.004)**† | 0.105 (0.004)† | 0.105 (0.003)† |
| ZDT1 d20 m2 | 1.98 (0.08) | 1.28 (0.2)† | **0.935 (0.1)**† | 1.89 (0.1) | 1.70 (0.2) | 1.12 (0.1)† | 1.21 (0.09)† |
| ZDT2 d20 m2 | 3.23 (0.08) | **1.56 (0.2)**† | 2.06 (0.2)† | 1.82 (0.2)† | 1.94 (0.2)† | 1.76 (0.2)† | 1.95 (0.2)† |
| *multiplicative noise* | | | | | | | |
| CarSideImpact | 0.104 (0.003) | 0.0995 (0.007) | 0.104 (0.005) | 0.0904 (0.006) | 0.0989 (0.002) | **0.0640 (0.005)**† | 0.104 (0.006) |
| DTLZ1 d15 m2 | 646 (30) | 430 (30) | 421 (30) | 513 (30) | 648 (20) | **245 (4)**† | 247 (3)† |
| DTLZ1 d15 m3 | 512 (30) | 402 (30) | 420 (20) | 422 (20) | 509 (30) | **212 (5)**† | 216 (5)† |
| DTLZ1 d20 m2 | 939 (30) | 744 (30) | 797 (20) | 784 (40) | 925 (30) | 350 (6)† | **341 (7)**† |
| DTLZ1 d20 m3 | 775 (20) | 633 (20) | 611 (30) | 702 (20) | 760 (20) | **302 (7)**† | 315 (7)† |
| DTLZ2 d15 m2 | 0.663 (0.02) | 0.573 (0.03)† | 0.557 (0.03)† | **0.465 (0.04)**† | 0.556 (0.02)† | 0.723 (0.05) | 0.520 (0.08)† |
| DTLZ2 d15 m3 | 0.692 (0.02) | 0.631 (0.03) | 0.620 (0.03) | **0.517 (0.03)**† | 0.548 (0.01)† | 0.748 (0.04) | 0.667 (0.06)† |
| DTLZ2 d20 m2 | 0.980 (0.02) | 0.774 (0.04) | 0.805 (0.04) | 0.681 (0.04) | 0.840 (0.03) | 0.947 (0.08) | **0.497 (0.1)**† |
| DTLZ2 d20 m3 | 0.982 (0.02) | 0.834 (0.03) | 0.876 (0.03) | **0.630 (0.05)**† | 0.703 (0.02)† | 1.07 (0.02) | 0.822 (0.1)† |
| DTLZ2 m5 | 0.643 (0.01) | 0.612 (0.02) | 0.619 (0.02) | **0.528 (0.03)**† | 0.560 (0.02)† | 0.735 (0.02) | 0.736 (0.02) |
| DTLZ7 m5 | 0.199 (0.009) | 0.0232 (0.005) | 0.0431 (0.009) | 0.0877 (0.01) | 0.0622 (0.01) | **4.67e-5 (3e-5)**† | 0.00603 (0.004) |
| Penicillin | 0.0698 (0.001)† | **0.0661 (0.002)**† | 0.0666 (0.003)† | 0.0741 (0.003)† | 0.0669 (0.003)† | 0.0808 (0.003) | 0.0857 (0.003) |
| VehicleSafety | 0.181 (0.009) | 0.0647 (0.02) | 0.0268 (0.01)† | 0.0265 (0.008)† | 0.0849 (0.01) | **0.0154 (1e-3)**† | 0.0375 (0.01)† |
| WFG6 m4 | 0.123 (0.002) | 0.125 (0.004) | 0.126 (0.004) | 0.123 (0.005) | 0.110 (0.005) | **0.0625 (0.002)**† | 0.0647 (0.004)† |
| ZDT1 d20 m2 | 2.01 (0.08) | 1.69 (0.2) | 1.67 (0.2) | 2.02 (0.1) | 2.06 (0.09) | **1.08 (0.1)**† | 1.15 (0.1)† |
| ZDT2 d20 m2 | 3.23 (0.08) | 2.55 (0.2) | 2.28 (0.3)† | 2.84 (0.2) | 3.19 (0.08) | **1.49 (0.3)**† | 2.17 (0.2)† |