

APPENDIX B

In this appendix, we provide detailed results with statistical analyses for RQ1, RQ2, and RQ4. First, for each problem and each time budget, we compare a pair of algorithms. Second, to compare the overall performance of the algorithms, we combine all objectives together by calculating average values of the objective functions (called *OFV*):

$$OFV = \frac{\sum_{i=1}^n Fitness_i}{n}$$

where n is the number of objectives for the prioritization problem, and $Fitness_i$ is the fitness value of the i th objective for the problem. Third, according to the guideline in [1], we used Hypervolume (*HV*) and Inverted Generational Distance (*IGD*) to compare the overall performance of multi-objective search algorithms. Last, we calculated *Rank* and *Confidence* (as described in Section 4.1.5) for group comparison. Table 1 shows the navigation of the results for RQ1, RQ2, and RQ4.

TABLE 1
Navigation of detailed results in the supplementary materials

Name	Index and Table Numbers
Use Case	AW1, AW2, AW3, AW4 and GS1 in Appendix B Sections B.1–B.5
RQ1	<ul style="list-style-type: none"> - Comparison with RS with the Vargha and Delaney statistics (\hat{A}_{12}) and Mann-Whitney U Test (p-value) in Table i ($i = 1 \dots 10$) for problem i - Multiple hypothesis testing with the Holm-Bonferroni method in Table 11
RQ2	<ul style="list-style-type: none"> - Group comparison with the Kruskal-Wallis test in Table 12 + $3(i - 1)(i = 1 \dots 10)$ for problem i - Pair comparison with the Vargha and Delaney statistics (\hat{A}_{12}) and Mann-Whitney U Test (p-value) in Table 13 + $3(i - 1)(i = 1 \dots 10)$ for problem i - Rank and Confidence in Table 14 + $3(i - 1)(i = 1 \dots 10)$ for problem i - Multiple hypothesis testing with the Holm-Bonferroni method in Table 42
RQ4	<ul style="list-style-type: none"> - Group comparison with the Kruskal-Wallis test in Table 43 for all the problems - Pair comparison using the Vargha and Delaney statistics (\hat{A}_{12}) and Mann-Whitney U Test (p-value) in Table 44 for all the problems - Multiple hypothesis testing with the Holm-Bonferroni method in Table 45

REFERENCES

- [1] S. Ali, P. Arcaini, D. Pradhan, S. A. Safdar, and T. Yue, “Quality indicators in search-based software engineering: An empirical evaluation,” *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 29, no. 2, pp. 1–29, 2020.