Algorithm Configuration Landscapes:
More Benign than Expected?

Supplementary Material

Yasha Pushak[1] and Holger Hoos[2,1]

[1] Department of Computer Science, The University of British Columbia, Vancouver, Canada
[2] LIACS, Universiteit Leiden, Leiden, The Netherlands

ypushak@cs.ubc.ca, hh@liacs.nl

A Introduction

This document contains some additional data and analysis collected for but not presented in the paper "Algorithm Configuration Landscapes: More Benign than Expected?" [1].

In Section B we show the additional 14 interesting parameter response slices for instance sets, in Section C we discuss the fitness-distance correlation (FDC) analysis of the parameter response slices, and in Section D we show the nine individual instance parameter response slice replicates we performed to verify that some individual instances do have highly rugged responses.

B Interesting Parameter Response Slices

We plot the 14 additional interesting instance set parameter response slices in Figures 1 and 2.

C Fitness-Distance Analysis

In the left of Figure 3 we plot the CDF of the FDC for the parameter response slices on the entire instance sets. To our surprise, the variance in the FDC across the bootstrap samples of the parameter slices is very high and the median FDC is relatively low for most parameters, e.g., 80% of the parameter responses have their median FDC less than 0.4. This large variance in the FDC tells us that the variation among the parameter responses on individual instances is relatively high, meaning that the aggregate response on the instance set, and hence the FDC, depends strongly on the instances in a particular bootstrap sample. However, we can also see that the FDC tends to be higher for parameters with interesting, non-flat responses. As a result, we believe that the relatively

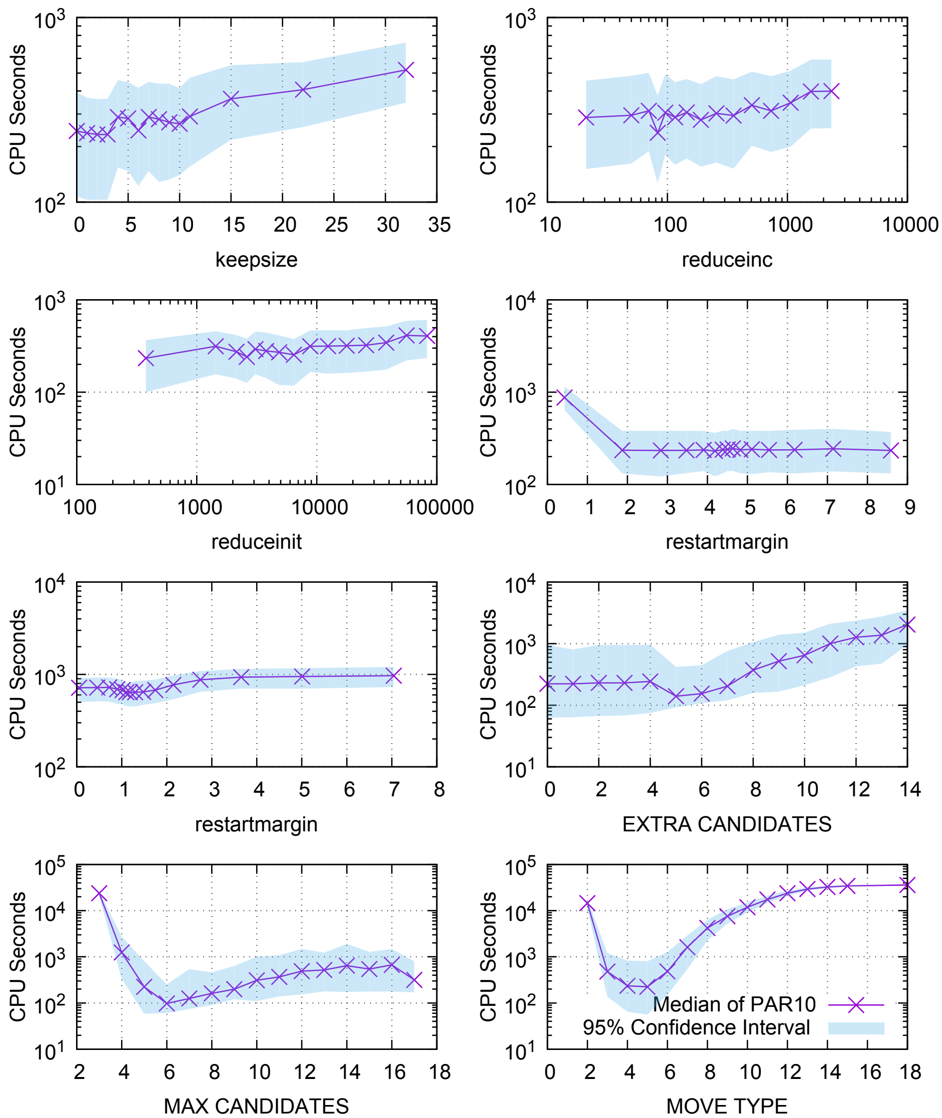

Figure 1. From left to right, top to bottom: CaDiCaL's `keepsize`, `reduceinc`, `reduceinit` and `restartmargin` on the circuit-fuzz instances and `restartmargin` on the BMC08 instances; and, LKH's `EXTRA_CANDIDATES`, `MAX_CANDIDATES` and `MOVE_TYPE` on the tsp-rue-1000-3000 instances.

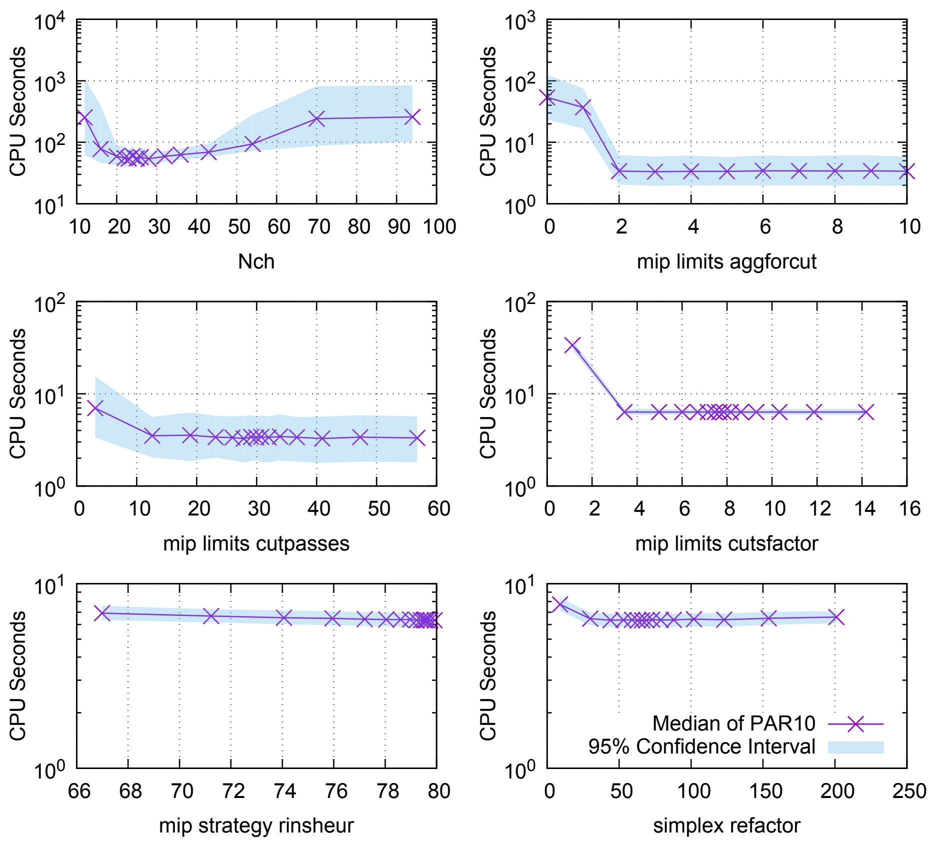

Figure 2. From left to right, top to bottom: EAX's `Nch` on the tsp-rue-1000-3000 instances; and CPLEX's `mip_limits_aggforcut` and `mip_limits_cutpasses` on the CLS instances and `mip_limits_cutsfactor`, `mip_strategy_rinsheur` and `simplex_refactor` on the Regions200 instances.

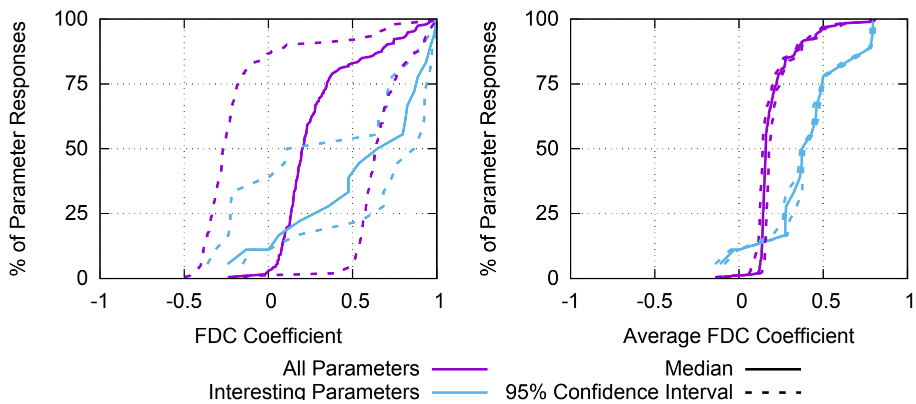

Figure 3. Left: the CDF of the median and 95% confidence interval of the FDC coefficients of the parameter response slices on the entire instance sets; right: for each parameter we took the average FDC over each of the individual instance parameter response slices and we show the resulting CDF of the average median and 95% confidence intervals. The confidence intervals are based on bootstrap sampling.

low FDC values are primarily due to relatively small fluctuations occurring in large, flat regions of the parameter responses (which tend to occur for most parameters).

Despite the fact that many of the individual instance parameter response slices appear to be uni-modal and convex, we see in the right of Figure 3 that the average FDC for many of the parameters is quite low, e.g., 80% of the parameter response slices have an average instance FDC less than 0.25. while some of the instance parameter response slices have negative FDCs, this does not necessarily indicate that they are highly rugged (consider $\frac{1}{x} + \log(x)$, which has very high values close to the right of the optimum near 2 and relatively low values far to the left of 2). However, given that the bootstrap confidence intervals for the FDCs are very small compared to those for the instance set parameter responses, we can infer that the ruggedness we observe on individual instances is more often truly reflective of the nature of the parameters' responses, rather than being simply due variation between independent runs of the algorithms. We can again see that the parameters with interesting responses (on the instance sets), tend to have individual instance responses with a higher FDC, again indicating that some of the low FDC scores we are seeing are due to relatively flat regions of the parameter responses.

D Parameter Response Slice Replicates

Through manual inspection, and in light of the large number of modes observed for some of the individual instance responses (see Figure 2 in the original paper), we know that we did obtain some very rugged individual instance parameter responses. To ensure that these were not spurious results, we performed replicates of nine of the responses. We plot the original and replicate parameter responses in Figures 4, 5 and 6. See the original paper for details on how the instances and parameters for the replicates were chosen.

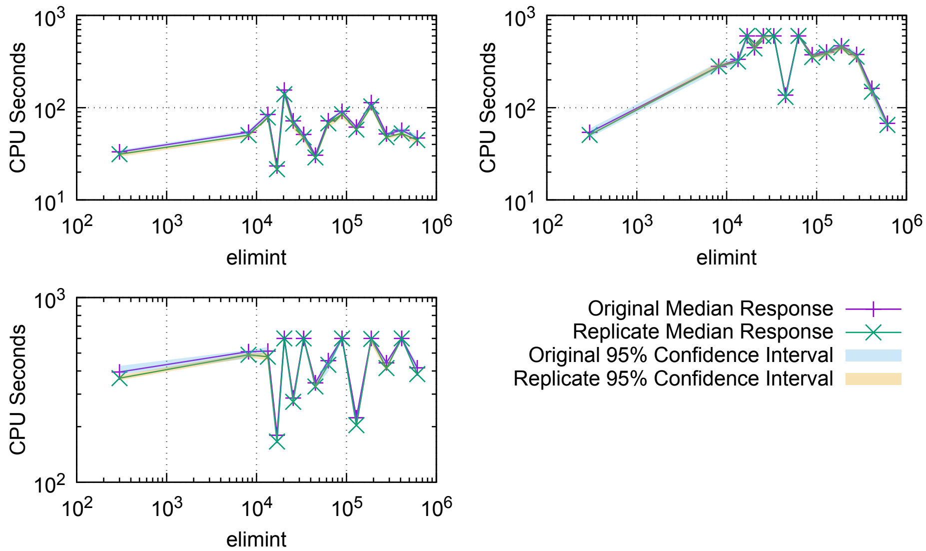

Figure 4. Replicates of rugged parameter response slices for CaDiCaL's `elimint` on three circuit-fuzz instance. From top to bottom, left to right: fuzz_100_634.cnf, fuzz_100_30719.cnf and fuzz_100_9873.cnf.

References

1. Pushak, Y., Hoos, H.H.: Algorithm configuration landscapes: More benign than expected? In: Proceedings of the 15th International Conference on Parallel Problem Solving from Nature (PPSN 2018). pp. 271–283 (2018)

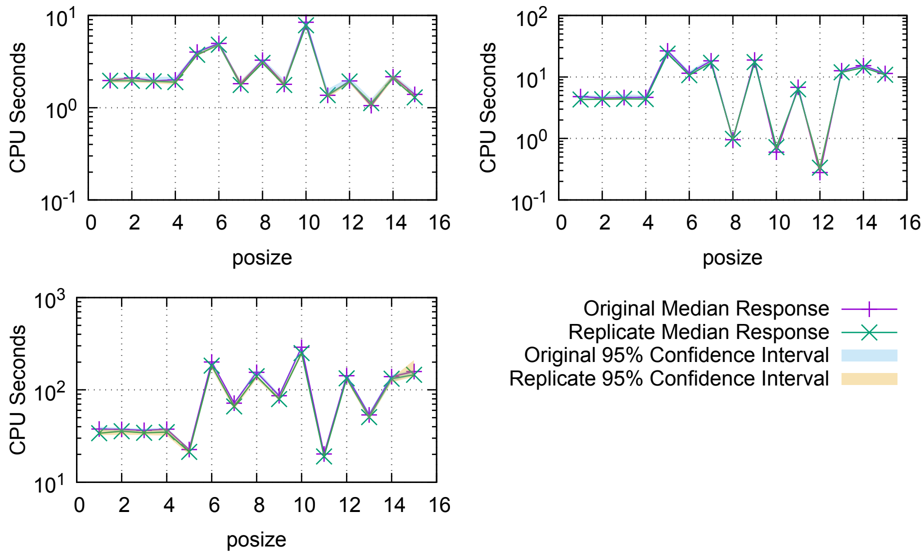

Figure 5. Replicates of rugged parameter response slices for CaDiCaL's `posize` on three circuit-fuzz instances. From top to bottom, left to right: fuzz_100_5082.cnf, fuzz_100_29685.cnf and fuzz_100_16079.cnf

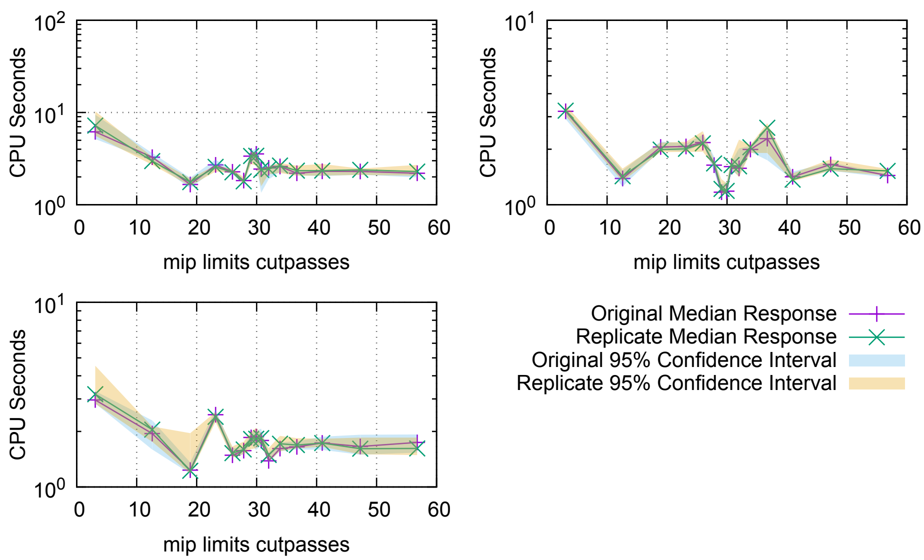

Figure 6. Replicates of rugged parameter response slices for CPLEX's `mip_limits_cutpasses` on three CLS instances. From top to bottom, left to right: cls.T90.C3.F200.S3.mps, cls.T90.C3.F1000.S5.mps and cls.T90.C5.F200.S5.mps.