We thank the reviewers for their helpful comments. We will correct the typos and improve the presentation according to their suggestions (such as including more details on quantitative semantics for STL).

1) Review 1: “I would expect many more benchmarks to showcase the strength of the approach across different domains.”

Response: In this work, our goal is to make a step towards a systematic combination of search algorithms which is agnostic to the search algorithms that are used. Indeed, we only exploit the main characteristics (exploitation vs exploration) of the algorithms. Given the heuristic nature of these search algorithms, it is very hard to derive a systematic exploitation of their specific functioning that works efficiently for a general problem. In the future, we plan to apply the approach to domains beyond automotive control.

2. Review 3: “The technique described in the ”iteratively updated initial point sampling distribution” seems to be very similar to simulated annealing….”

Response: We assume that the mentioned similarity refers to population-based Simulated Annealing (SA) with restarts, which essentially involves running in parallel a number of Simulated Annealing and restarting from some previous best solutions (rather than continuing with the current solutions). A restart is generally triggered either after some predefined number of steps or based on the current energy.

We agree that there is an overlap with the SA method with restarts, our restart strategy is close to the idea of population-based approaches such as CMA-ES. However, our approach differs from these algorithms in the way a restart is triggered. Besides a predefined number of steps, our approach can trigger a restart based on the detection of a possible blocking situations around a local optimum, by monitoring the evolution of the coverage (of the visited candidate points) and the robustness (that is the value of the objective function). Our approach chooses between exploitation and exploration based on this monitoring, and for this reason, the algorithm is “broken out as two separate phases”, as stated in the review. Also, our restart strategy can be used with any search algorithm and is not tied to specific variables of an algorithm (such as energy in SA).

3) Review 3: “interested in seeing other global search algorithms employed when the local search gets stuck…”

Response: We agree that more experimental results would be ideal. We note that we do provide a comparison of our meta-heuristic approach to (pseudo) random sampling-based optimization. Additionally comparing random sampling to the “iteratively updated initial point sampling”, which would correspond to an evaluation of one component of our overall method, would have been informative, but our goal was to prioritize evaluation of the overall approach.

4) Review 2: “include this in their related work (and comparisons?): Two-layered falsification…”

Response: We will contrast our method with this work (referred to herein as [Z18]), which has a common flavor with our previous work (ref [5] in [Z18]), where classification is used to partition the input space and sampling is based on balancing the current potentials of exploitation and exploration. [Z18] also uses a similar balancing mechanism by biased random sampling. Our current submission uses a balancing mechanism that is more deterministic based an explicit coverage measure. Another important difference is that [Z18] (as most other related work) considers fixed time discretization (more concretely, equal-step time discretization), while our submitted work considers variable step discretization together with a notion of timed pattern coverage (see the next paragraph). While a direct comparison with that work would be ideal, we remark that for the Automatic Transmission benchmark, [Z18] considers 5 control points for the two input signals, our submitted work considers 7 and 3 control points but with smaller value ranges, i.e. [35 100] and [0 40 ] vs [0, 100] and [0, 325] in [Z18]. For a similar property (S3) considered by [Z18], the computation times are “comparable” by an order of magnitude (950s in average for 100% success rate in our work, and 666s for 80% success rate in [Z18]).

5) Finally, we would like to emphasize that, in addition to using coverage of values of input signals to guide our search algorithm, our work presents the novel idea of using timed pattern coverage for generating input signals satisfying temporal constraints (described by timed automata). This is in contrast to existing optimization-based falsification approaches, which require predefined time patterns (or time discretizations) of input signals. By using the timed pattern coverage based input signal generation, our approach is able to focus on only input signals that satisfy the given temporal constraints; moreover, using a measure of coverage for this set of input signals, we can quantify the confidence in the test result.