

As there are many ways to combine each configuration, we select a few configurations by examining them through simulations.

The set of model-configuration contains 54 elements. We choose to run each model configuration for 100 evolution cycles. For all model configurations, we use the same four factual process instances randomly sampled from the test set. We ensure that the outcomes of these factuals are evenly divided. We decide to limit the population size to a maximum of 1000 counterfactuals. Within each evolutionary cycle, we generate 100 new offspring. We keep the mutation rate at 0.01 for each mutation type. Hence, across all cases that are mutated, the algorithm deletes, inserts, and changes 1% of events per cycle. We collect the mean viability and its components across the iterative cycles of the model.

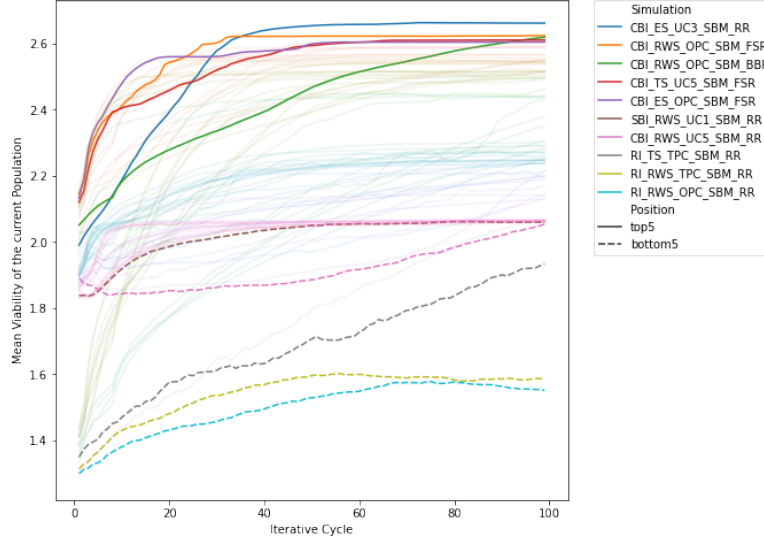


Figure 1: This figure shows the average viability of the five best and worst model configurations. The x-axis shows how the viability evolves for each evolutionary cycle.

Figure 1 shows the bottom and top-5 model configurations based on the viability after the final iterative cycle. We also show how the viability evolves for each iteration. The results reveal a couple of patterns. First, all top-5 algorithms use either *Case-Based-Initiator* as initiation operation. In contrast, the bottom-5 use *Random-Initiator* as initialisation. Hence, the initialisation appears to be majorly important for the algorithm. The complete table of results is in ??.

In Figure 2, we see the effects of each operator type.

Starting with some commonalities, across operator-type and measure, the figure shows that the initiator heavily determines the starting point for each measure. For instance, the *Random-Initiator* starts well below the other

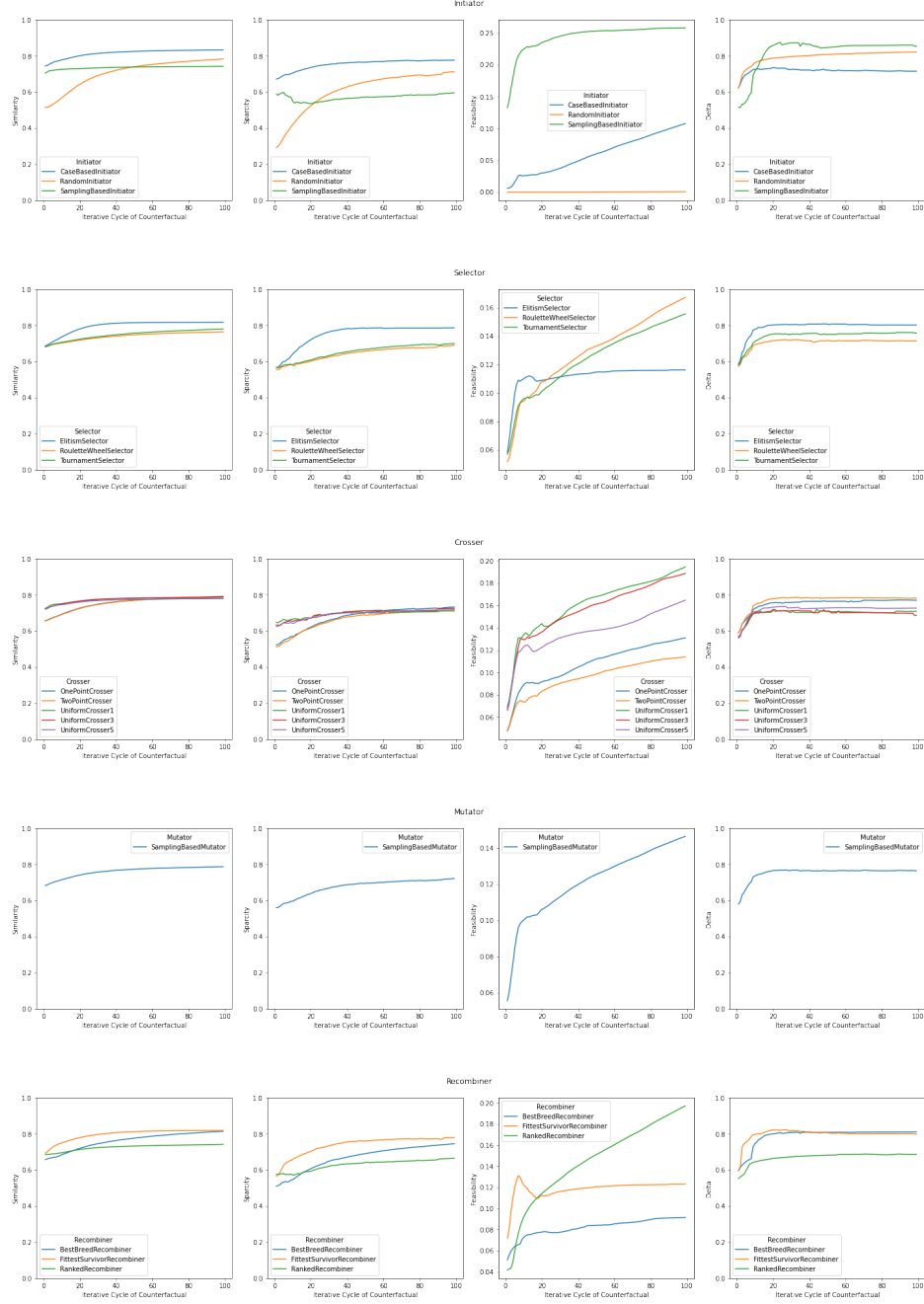


Figure 2: The evolution of each viability measure over the entire span of iterative cycles. Each figure adjust the respective operator type by taking the average over all other operator types.

initiator operations regarding sparsity and similarity. Similarly, most of the *RI-x* model configurations begin at much lower viability than the other model configurations. This pattern is evident in Figure 1.

Another noteworthy general observation is the delta measure. Here, we see a movement towards the highest possible delta value for each operator type. Hence, most configurations are capable of changing the source outcome to the desired outcome.

Regarding feasibility, Figure 2 shows an increase for most operators. This is especially true if the operator has a random component or if it optimises for feasibility. Similar holds for recombination with *Fittest-Survivor-Recombiner*. The feasibility has not reached convergence yet. In many cases, the feasibility monotonously increases.

Among the top-5 *CBI-ES-OPC-SBM-FSR* grows the fastest in terms of viability and reaches convergence the earliest. On the opposite side, we find *CBI-RWS-OPC-SBM-BBR* to have the slowest growth. However, it is also the only one not reaching convergence at that point.

When it comes to the crossing operation, the results indicate that the differences between *One-Point-Crosser* and *Two-Point-Crosser* are inconclusive for all viability measures except feasibility. One can explain that by noting that both operations are very similar. However, cutting the sequence only once produces fewer impossible sequences for the child sequences.