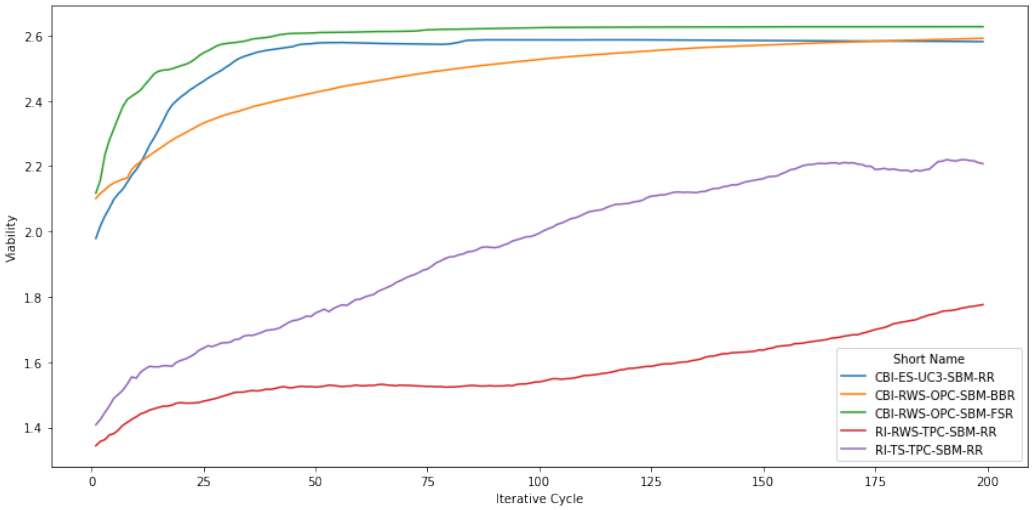

Figure 1: This figure shows the viability of across the iteration cycles.

In Figure 1, we see a general increase in viability for each termination point. It shows, that increasing the termination point also yields better results at the end of the generation process. We see that **[CBI-ES-UC3-SBM-RR]** returns the best results in the shortest time span. The model converges after roughly 50 iterative cycles. **[CBI-RWS-OPC-SBM-BBR]** appears to have not reached convergence. The randomly initiated models have not reached convergence as well. However, they remain far below models that use a more sophisticated method to initialize their population.

Figure 2 shows a decomposed view on how the viability measure evolves. Furthermore, we show the average amount of events within a generated counterfactual. In terms of similarity and sparcity all models behave similar. This is no surprise as both measures are ineherently interlinked. We see that the randomly initiated models (RI-x) decrease the amount of events they generate. Case-based initiated models appear to slightly gain more events. Although, **[CBI-RWS-OPC-SBN-BBR]** appears that reaches its saturation point significantly later (**[100]**th cycle). Interestingly, the **[CBI-RWS-OPC-SBM-BBR]** model struggles to maintain feasibility and collapses to near 0 after the **[100th]** iterative cycle. Another surprise is the steep ascention of only model that uses tournament selection (RI-TS-TPC-SBM-RR) towards the end of the generation process. The model even overtakes the model that leads the model-configurations in terms of **[viability]**. Furthermore, we see that CBI-ES-UC-SBM-RR has the highest feasibility among all models. However, it also quickly converges after 50 iterative cycles.

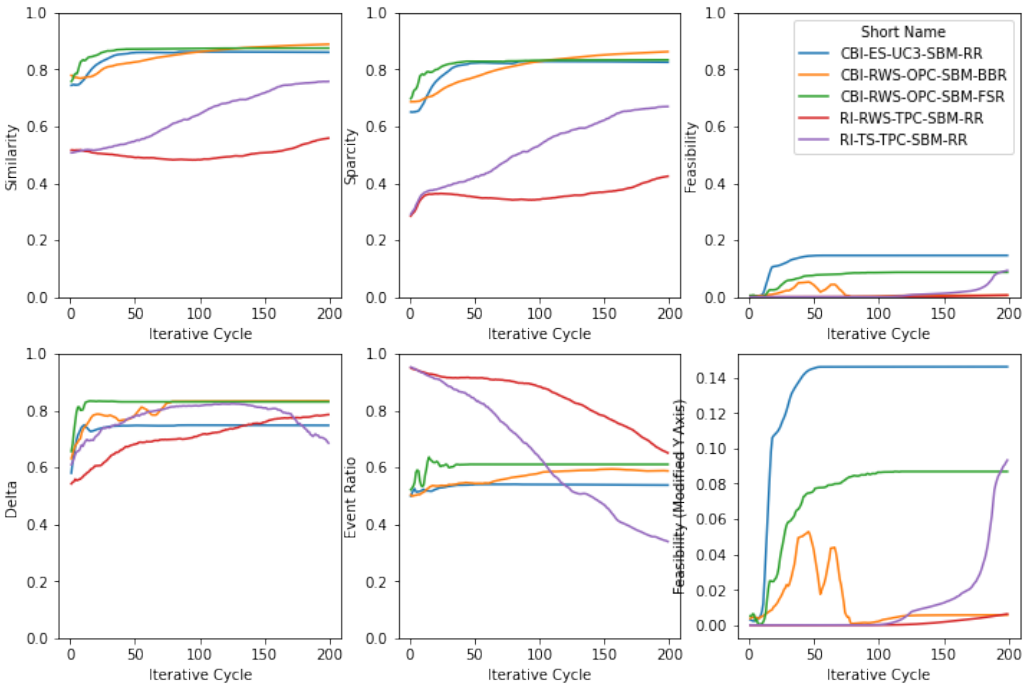

Figure 2: This figure shows the remaining measure components. Additionally, we show the ratio of events within the population. We also show a magnified version of the feasibility measure.