COMP 6666 Fall 2020 Assignment 2c

Dennis Brown, dgb0028@auburn.edu

November 29, 2020

1 Introduction

This assignment introduces Competitive Coevolutionary Genetic Programming (GP) to the game of Pac-Man.

1.1 Deliverables

- GREEN 1: Please see README.md in the repository for details on these deliverables.
- GREEN 2: Please see README.md in the repository for details on these deliverables.
- GREEN 3: Section 2 of this report fulfills the GREEN 3 deliverable.
- GREEN 4: Please see README.md in the repository for details on these deliverables.
- YELLOW 1: Section 3 of this report fulfills the YELLOW 1 deliverable.

2 GREEN 3 Report: Basic Investigation

2.1 Methodology

In this Competitive Coevolutionary Genetic Programming (CCEGP[1])-based search:

- Compared to the previous assignment, there are now two populations, Pac-Man and the Ghosts.
- The basic components of GP are carried over from the previous assignment, with minor alterations, and applied to both populations:
  - Individuals in both populations are encoded as binary expression trees with nodes that are either functions (interior nodes) that perform one of five operations (+, -, *, /, RAND) on the values of each child node, or inputs (leaf nodes) that are one of six values (five metrics based on game state or a constant value).
  - Initialization of both populations is Ramped half-and-half with a depth limit of D_{max} as defined in [2]. It randomly chooses (with equal probability) between the "grow" and "full" methods.
  - Parent selection for both populations is implemented as Fitness Proportional Selection (as implemented in previous assignments) or Overselection as defined in [2]; Overselection selects from the top 32% with 80% chance and from the rest with 20% chance.
  - Recombination and mutation for both populations are implemented as described in [2]. The code walks through the selected parents. With a configurable chance of mutation (set at 5%), it chooses to mutate a copy of the current parent as the next offspring, otherwise it recombines the current and next parents to make two offspring.
    - Mutation randomly picks a node in the tree and replaces it with a re-grown sub-tree using the initialization capability in "grow" mode.
    - Recombination clones each parent as offspring, then randomly picks a node in each offspring and swaps them. Trees are allowed to exceed D_{max} (code to prevent that situation is implemented but disabled) with the thought that parsimony pressure will keep excessive trees in check.
  - Survivor selection for both populations is implemented as Truncation or k-Tournament (as implemented in previous assignments).

[1] The author didn't find a common acronym for Competitive Coevolutionary Genetic Programming and arbitrarily decided to use *CCEGP*.

  - Parsimony pressure for both populations is implemented at the end of each game/evaluation by calculating fitness by reducing the game score by either the size or height of the tree multiplied by the parsimony pressure penalty coefficient.

- Evaluation is primarily where CCEGP differs from GP. In this implementation, during every generation, fitness is calculated by playing a game between an individual in the Pac population and an individual in the Ghost population. Koza provides a generic co-evolution flowchart in Chapter 16 of his book[3], and states "In co-evolution, the relative fitness of a particular strategy in a particular population is the average of the payoffs that the strategy receives when it is played against fitness cases consisting of each strategy in the opposing population of strategies." In picking which individuals from each population play each other, one interpretation is that every individual plays every other individual; i.e. each member of the population has a unique strategy (thus a unique fitness case). As will be explained in Section 2.2, this "Individual vs All: approach was investigated first and computational limitations restricted population sizes in deleterious ways. A second approach ("Individual vs Individual") was the investigated in which individuals from the opposing populations are randomly paired, and as directed by the assignment, if an individual is used more than once, its fitness is the average of the games it played. NOTE: This work has a flaw in that, in each generation, the offspring only play against each other to calculate fitness (then survival selection proceeds as normal looking at population + offspring). In hindsight a better approach would be to play each offspring against an individual randomly sampled from the entire population. However, time constraints prevent remedying this situation.

2.2 Experimental Setup

The "baseline" GP-specific settings in each configuration file are as follows (variations will be described as needed). In the case of settings specified per-population, they are identical for both populations (Pac-Man and Ghost) unless otherwise pointed out:

- 30 runs of 2000 evaluations each (termination at 2000 evaluations)
- timer-initialized random seed
- pill density, fruit spawn probability, fruit score, time multiplier = 50%, 1%, 10, 2
- $D_{max} = 7$ for initialization
- $D_{max} = 9$ post-initialization
- Parent selection: Overselection of top 32%
- Probability of mutation: 0.05
- Survival strategy: $\mu + \lambda$ with $\mu = 100$, $\lambda = 50$ (only the "plus" survival strategy is implemented)
- Survival selection: Truncation (k-Tournament also implemented)
- Parsimony: Based on size (rather than depth) with a parsimony pressure penalty coefficient of 0.5

This investigation set up three configuration files each for two approaches to CCEGP:

1. **"Individual vs All" Evaluation Approach:** As described, during initialization and the amongst offspring in every generation, every member of each population plays against every other member of the opposing population and fitnesses are averaged for each individual. Each of these configurations builds upon the "baseline" configuration above with these exceptions: Due to the sheer number of evaluations required in each generation, the population sizes were kept small at with $\mu = 10$, $\lambda = 5$; and in an early attempt to reduce computation time, D_{max} was made smaller at 5 for initialization and 7 for post-initialization. The further unique characteristics of each configuration are listed below, and apply to both populations.
   - config1: Fitness Proportional parent selection. Justification: see if parent selection method significantly changes the outcome.
   - config2: Overselection (top = 32%) for parent selection (this is essentially the baseline configuration for this approach). Justification: arbitrarily chosen as a starting point. ['config1' and 'config2' should be swapped, in hindsight]
   - config3: Fitness Proportional parent selection; k-tournament (k=5) survival selection. Justification: see if survival selection method significantly changes the outcome.

2. **"Individual vs Individual" Evaluation Approach:** As described, during initialization and the amongst offspring in every generation, every member of each population plays against a single randomly-sampled member of the opposing population and if a member plays multiple times, fitnesses are averaged. Each of these configurations builds upon the "baseline" configuration above, except for the unique characteristics of each configuration listed below, and they apply to both populations unless otherwise specified.

   - config1: (no changes–this is the baseline configuration). Justification: arbitrarily chosen as a starting point.
   - config2: Fitness Proportional parent selection; k-tournament (k=10) survival selection. Justification: see if decreasing parent selection pressure but increasing survival pressure makes a difference.
   - config3: Ghost population was set to $\mu = 20$, $\lambda = 10$. Justification: See if a very small ghost population, which makes each Ghost play against several Pacs, makes a difference.

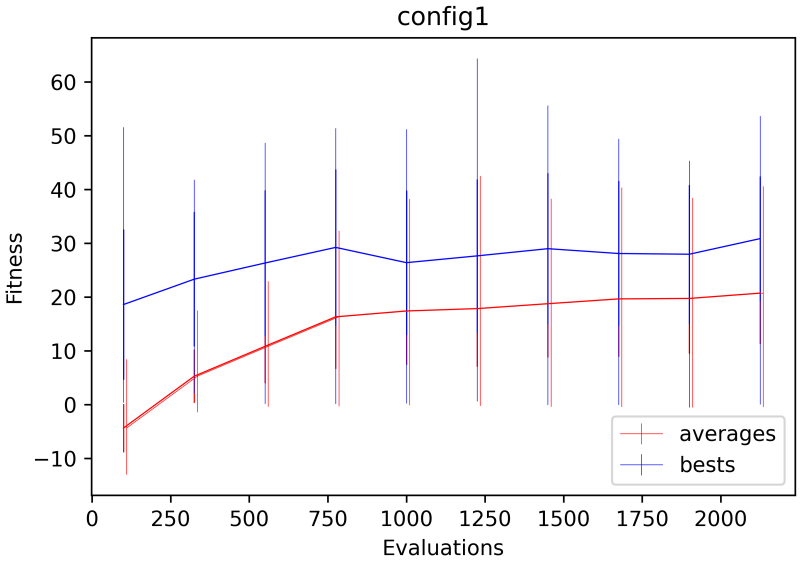
2.3 Results

NOTE: The statistical analysis for any comparison of two alternatives in this report consisted of performing an F-Test, determining equal or unequal variances, then performing a two-tailed t-Test. In all cases, $\alpha = 0.5$ and the null hypothesis is that the two samples are NOT statistically different.

2.3.1 Individual vs. All Approach

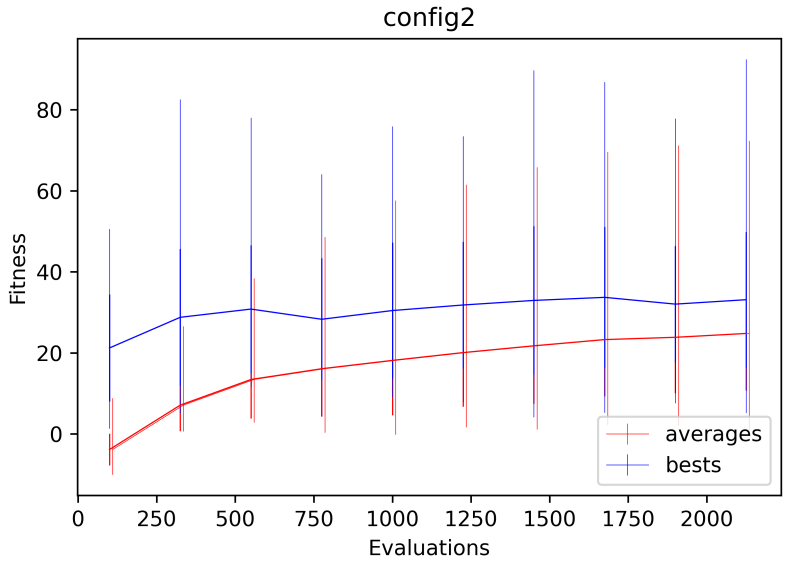
Average and Best Fitness (averaged over 30 runs) vs Evaluations for config1 are shown in Figure 1.

Figure 1: Average Averages and Average Bests for config1



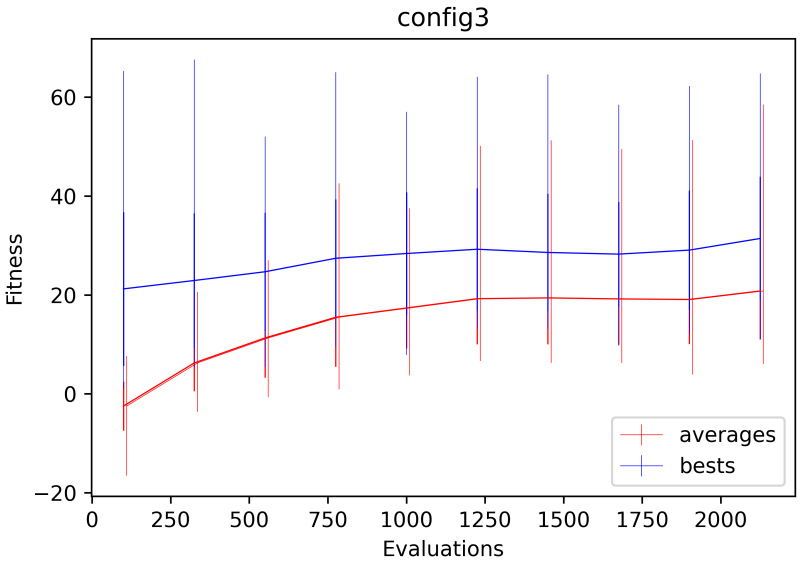
Average and Best Fitness (averaged over 30 runs) vs Evaluations for config2 are shown in Figure 2.

Figure 2: Average Averages and Average Bests for config2



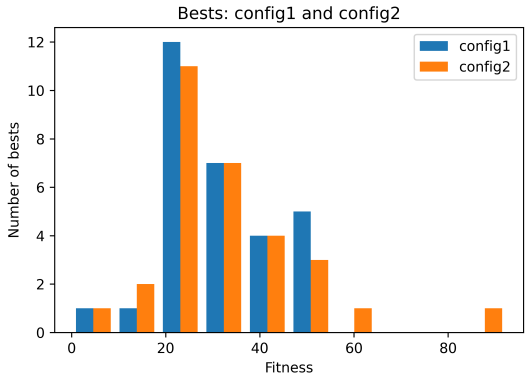
Average and Best Fitness (averaged over 30 runs) vs Evaluations for config3 are shown in Figure 9.

Figure 3: Average Averages and Average Bests for config3



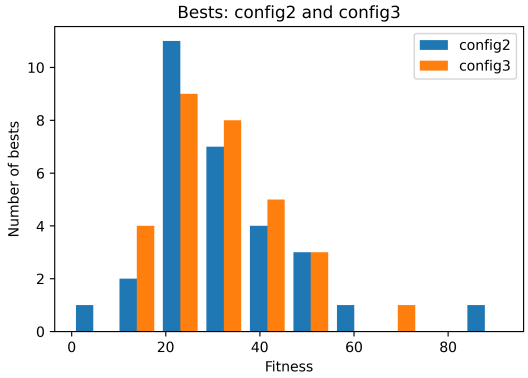
config1 vs config2: There is no statistically significant difference in the outcomes of these two variations. See the histogram in Figure 4 in this section and, in the Appendix, Table 1 for F- and t-test values.

Figure 4: config1 vs. config2 – Best values over 30 runs



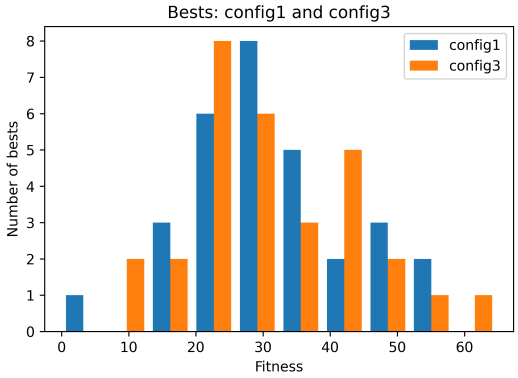
config2 vs config3: There is no statistically significant difference in the outcomes of these two variations. See the histogram in Figure 5 in this section and, in the Appendix, Table 3 for F- and t-test values.

Figure 5: config2 vs. config3 – Best values over 30 runs



config1 vs config3: There is no statistically significant difference in the outcomes of these two variations. See the histogram in Figure 6 in this section and, in the Appendix, Table 5 for F- and t-test values.

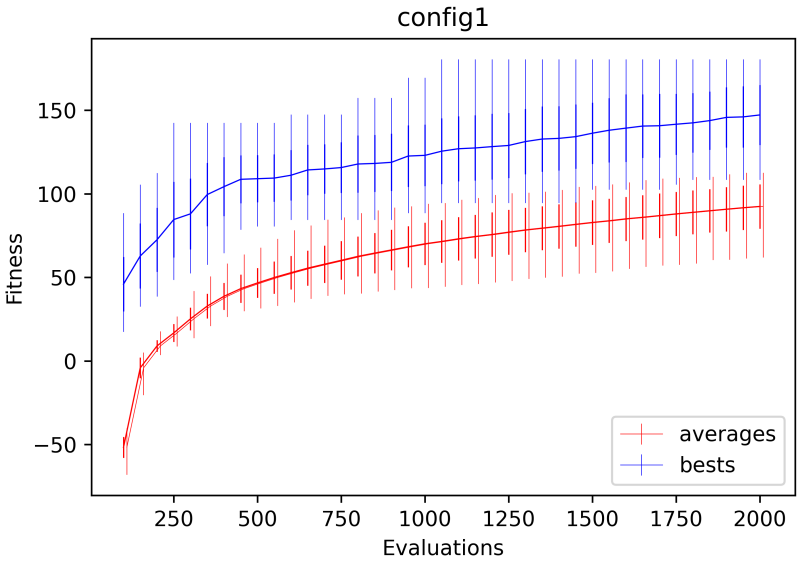
Figure 6: config1 vs. config3 – Best values over 30 runs



2.3.2 Individual vs. Individual Approach

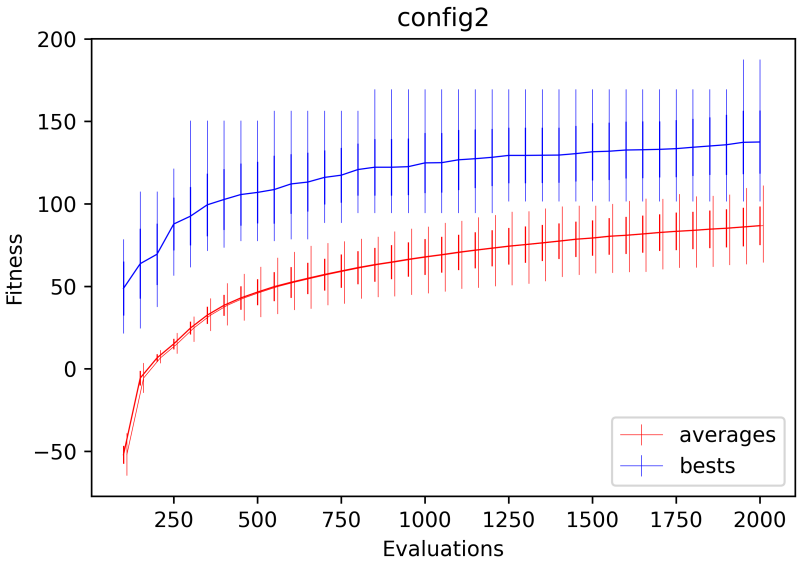
Average and Best Fitness (averaged over 30 runs) vs Evaluations for config1 are shown in Figure 1.

Figure 7: Average Averages and Average Bests for config1



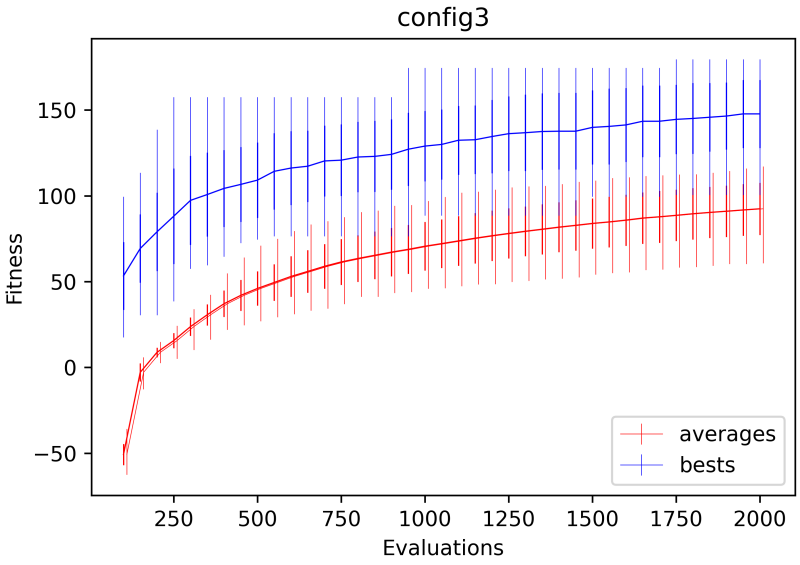
Average and Best Fitness (averaged over 30 runs) vs Evaluations for config2 are shown in Figure 8.

Figure 8: Average Averages and Average Bests for config2



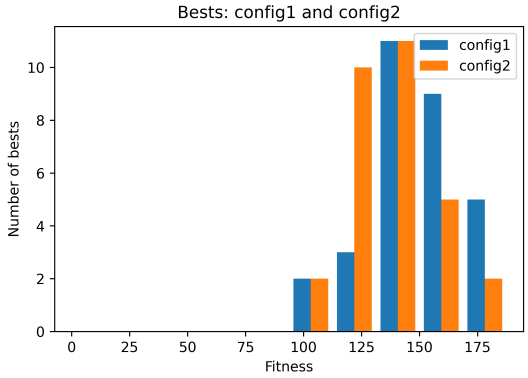
Average and Best Fitness (averaged over 30 runs) vs Evaluations for config3 are shown in Figure 9.

Figure 9: Average Averages and Average Bests for config3



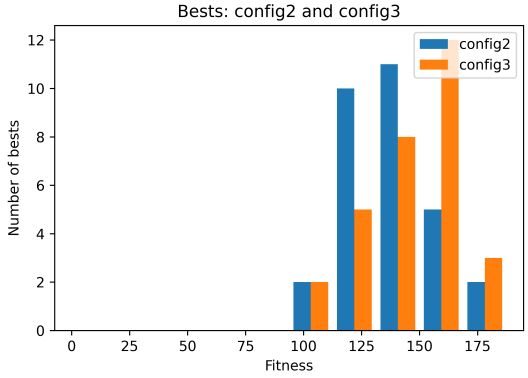
config1 vs config2: There is no statistically significant difference in the outcomes of these two variations. See the histogram in Figure 10 in this section and, in the Appendix, Table 7 for F- and t-test values.

Figure 10: config1 vs. config2 – Best values over 30 runs



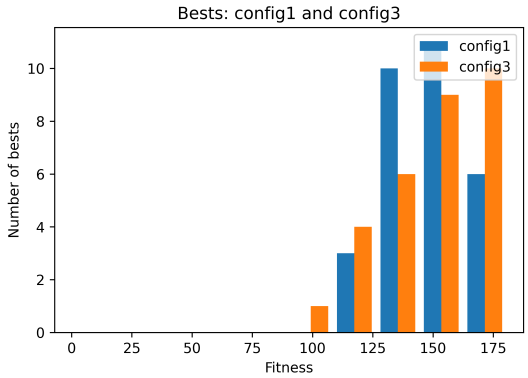
config2 vs config3: There is no statistically significant difference in the outcomes of these two variations. See the histogram in Figure 11 in this section and, in the Appendix, Table 9 for F- and t-test values.

Figure 11: config2 vs. config3 – Best values over 30 runs



config1 vs config3: There is no statistically significant difference in the outcomes of these two variations. See the histogram in Figure 12 in this section and, in the Appendix, Table 11 for F- and t-test values.

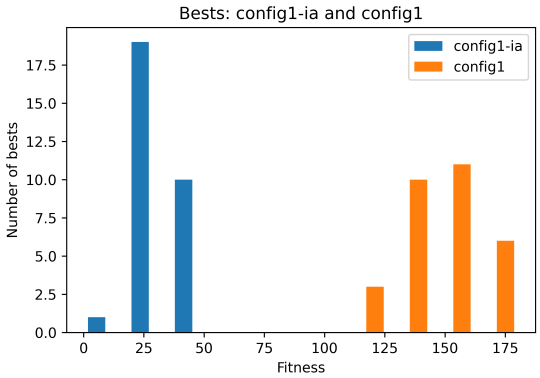
Figure 12: config1 vs. config3 – Best values over 30 runs



2.3.3 "Individual vs. All" Approach vs "Individual vs. Individual" Approach

Finally, we consider differences between "Individual vs All" and "Individual vs Individual" approaches. Since all "Individual vs All" are not statistically different, and all "Individual vs Individual" are not statistically different, we only compare config1 of each approach. In this case, "Individual vs Individual" is very significantly better than "Individual vs All" – making some mental leaps that this is even a meaningful comparison. See the histogram in Figure 13 in this section and, in the Appendix, Table 13 for F- and t-test values.

Figure 13: config1 vs. config1 – Best values over 30 runs



2.4 Discussion

These results of comparisons within each approach are largely inconclusive, and additionally, they are based on directly comparing absolute fitnesses of only the Pac-Man population (since that was all that was logged, per the assignment instructions; note that the author is not deflecting fault, just explaining circumstances). Changing the parent and survival selection methods, and even drastically changing the Ghost population size, did not reveal statistically significant differences within each of the two approaches.

The author then compared the first configuration of each of the approaches, which shows a very significant difference in performance; that is not surprising given that "Individual vs All" had much smaller populations and many fewer generations than "Individual vs Individual" and thus much poorer conditions for evolution. This comparison is only meaningful if one considers the evaluation limit of 2000 a control in that "Individual vs Individual" may be a better way to more effectively use those 2000 evaluations. In absence of evaluation limits, this comparison of approaches is not conclusive.

The author also created a Current Individual vs Ancestral Opponent (CIAO) plots for all 30 runs for each of these configurations; they can be viewed in the appendix in sections B.1 and B.2. The majority of these plots

show mediocre stability, with a few disengagements and occasional cycling. None clearly show "good evolution" as defined by Cliff and Miller[1].

2.5 Conclusion

The author has failed to characterize the sensitivity of the final local best to a specific parameter, as directed by the assignment. The author did find that, given a strict evaluation budget, the Individual vs Individual approach gives better results than Individual vs All, assuming one makes the leap of faith of comparing raw Pac-Man fitness values across those experiments.

3 YELLOW 1 Report Extension: Coevolutionary Cycling Investigation

3.1 Methodology

This work extends that described in Section 2. The same methodology was followed, but only the "Individual vs Indivdual" approach was used, and the runs were reduced from 30 to 10 since we are more concerned about viewing a reasonable number of CIAO plots rather than performing statistical analysis.

3.2 Experimental Setup

Two configurations were tested. They are based on the baseline of config1 for the "Individual vs Individual" approach:

- config4cyc: Pac population was set to $\mu = 10$, $\lambda = 5$. Ghost D_{max} was reduced to 2. Ghost parent selection was overselection of top 100% and survival selection was k-Tournament with k=1. Justification: See if a very small Pac population, which makes each Pac play against several Ghosts, and greatly decreasing pressure on the Ghost population, induces cycling.
- config5cyc: Pac population was reset to $\mu = 100$, $\lambda = 50$. Ghost population was reduced to $\mu = 1$, $\lambda = 1$ and D_{max} was reduced to 1. Ghost parent selection was overselection of top 100% and survival selection was k-Tournament with k=1. Justification: See if a normally-sized Pac population and greatly decreasing pressure on the Ghost population (practially random), induces cycling.

3.3 Results

As a sanity check, here are the Average and Best Fitness (averaged over 30 runs) vs Evaluations for config4cyc and config5cyc:

Figure 14: Average Averages and Average Bests for config4cyc

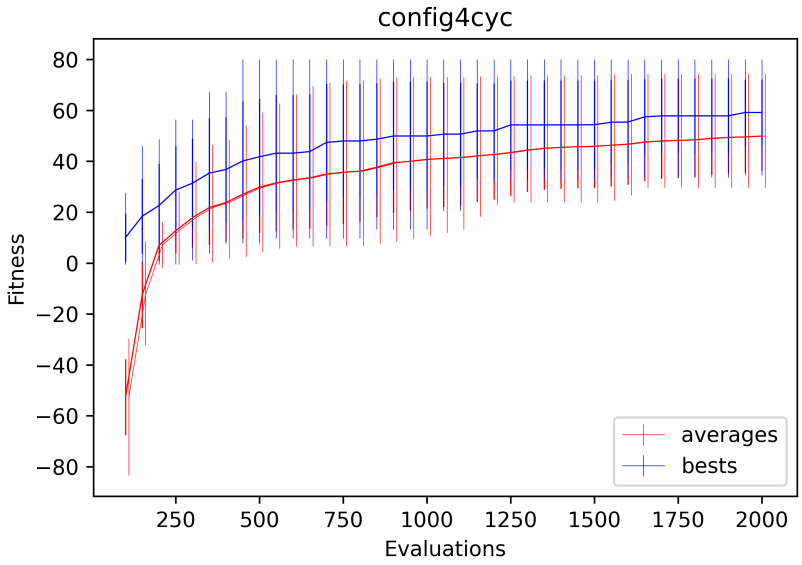
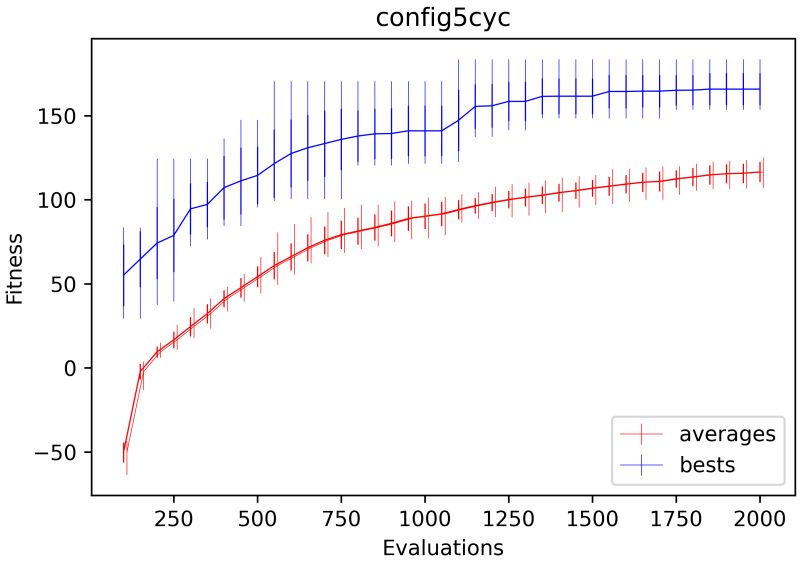


Figure 15: Average Averages and Average Bests for config5cyc



The CIAO plots for all 10 runs of each configuration are available in the appendix in Section B.3.

Also, the usual comparisons with the histogram, F-Test, and t-Test were performed, even though they are not the point of this investigation; if curious, the reader can find them in section A.2.

3.4 Discussion

As with the basic investigation configurations, most of the CIAO plots showed mediocre stability.

For config4cyc, 2 of 10 runs (#5 and #10) show signs of cycling; compare that to "Individual vs Individual" config1 with 5 of 30 runs (#4, #16, #19, #26, and #28), confgi2 with 2 runs (#2 and #14), and config3 with 2 runs (#12 and #25) showing similar signs of cycling (granted, these are very subjective observations). Without attempting a formal analysis, the author claims config4cyc doesn't exhibit significantly more cycling than configurations that were not specifically developed to exhibit cycling.

Figure 16: Signs of cycling for config4cyc: Run 5

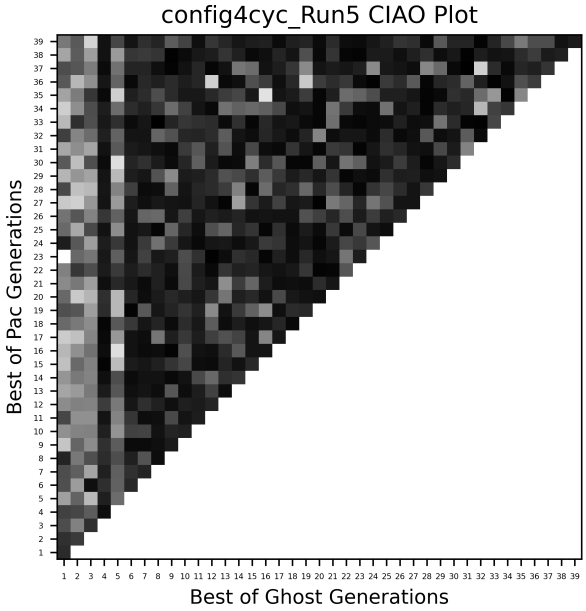
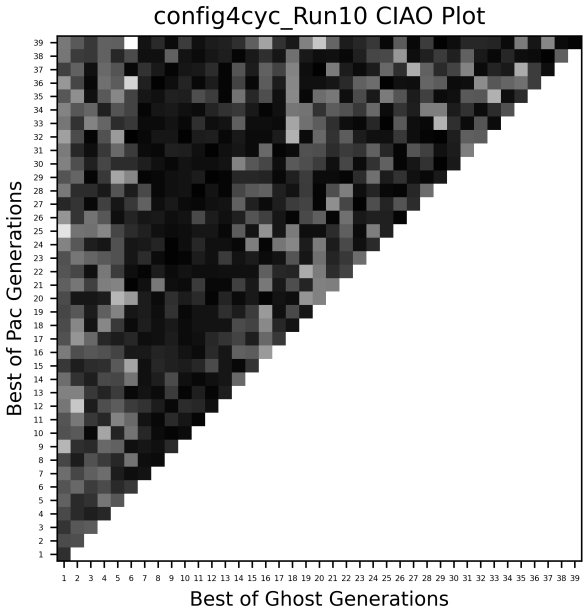


Figure 17: Signs of cycling for config4cyc: Run 10



For config5cyc, 2 of 10 runs (#1 and #3) show signs of cycling, and 2 more runs show weak signs of cycling (#4 and #5). So perhaps this configuration shows more cycling per run than the other configurations, but if so it's not a strong case.

Figure 18: Signs of cycling for config5cyc: Run 1

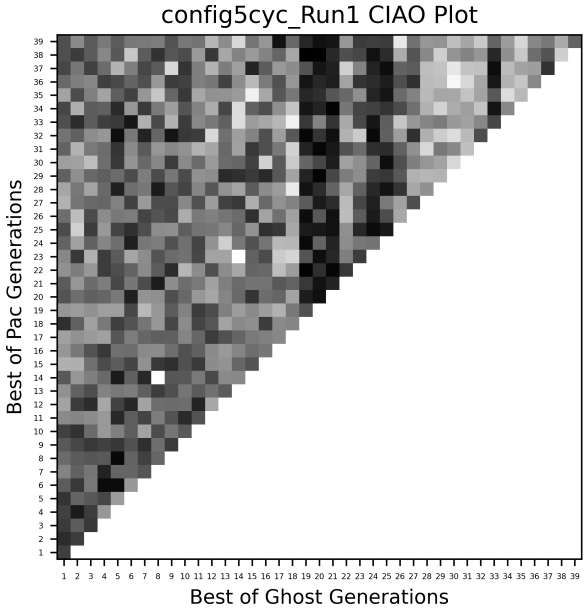
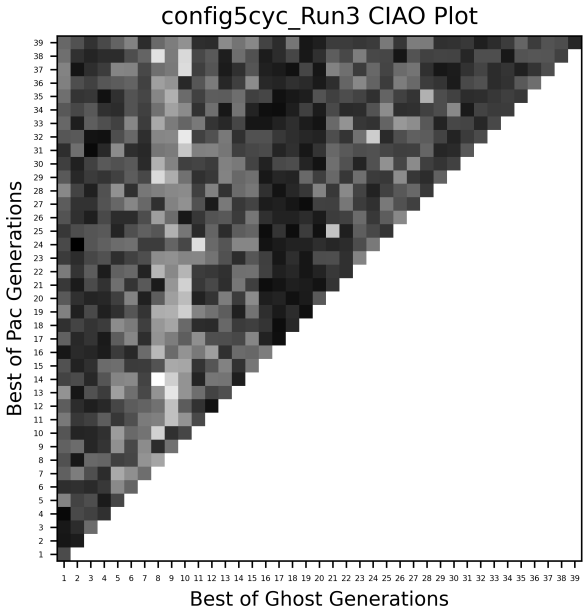


Figure 19: Signs of cycling for config5cyc: Run 3



3.5 Conclusion

The author attempted to induce cycling by introducing more randomness into the Ghost population, which in theory should have limited evolutionary memory, which is a condition that can cause cycling per Cliff and Miller[1]. However, the resulting number of runs showing clear signs of cycling did not seem to be significantly greater (without attempting a formal analysis) than the baseline configurations of the Basic (Green) investigation of this assignment.

References

[1] D. Cliff and G. Miller. Tracking the red queen: Measurements of adaptive progress in co-evolutionary simulations. In *ECAL*, 1995.

[2] A. E. Eiben and James E. Smith. *Introduction to Evolutionary Computing*. Springer Publishing Company, Incorporated, 2nd edition, 2015.

[3] John R. Koza. *Genetic programming - on the programming of computers by means of natural selection*. Complex adaptive systems. MIT Press, 1993.

A Appendix: F- and t-test tables

A.1 Basic Investigation

A.1.1 Individual vs. All Approach, config1 vs config2

Table 1: F-Test for config1 vs. config2 with $\alpha = 0.05$

	config1	config2
Mean	30.87555555555555	33.12888888888889
Variance	139.15784929757342	289.7914738186463
Observations	30	30
df	29	29
F	0.4801999433035751	
P(F≤f) one-tail	0.026368018399803 23	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 2: t-Test for config1 vs. config2 with Unequal Variances

	config1	config2
Mean	30.87555555555555	33.12888888888889
Variance	139.15784929757342	289.7914738186463
Observations	30	30
df	58	29
t Stat	-0.5959132955348058	
P(T≤t) two-tail	0.5538366208912282	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

A.1.2 Individual vs. All Approach, config2 vs config3

Table 3: F-Test for config2 vs. config3 with $\alpha = 0.05$

	config2	config3
Mean	33.12888888888889	31.397777777777776
Variance	289.7914738186463	162.0026768837803
Observations	30	30
df	29	29
F	1.788806699944476	
P(F≤f) one-tail	0.938437541554462	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 4: t-Test for config2 vs. config3 with Unequal Variances

	config2	config3
Mean	33.12888888888889	31.397777777777776
Variance	289.7914738186463	162.0026768837803
Observations	30	30
df	58	29
t Stat	0.4460825855490299	
P(T≤t) two-tail	0.6573310117234714	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

A.1.3 Individual vs. All Approach, config1 vs config3

Table 5: F-Test for config1 vs. config3 with $\alpha = 0.05$

	config1	config3
Mean	30.87555555555555	31.397777777777776
Variance	139.15784929757342	162.0026768837803
Observations	30	30
df	29	29
F	0.8589848758943927	
P(F≤f) one-tail	0.34252657099842354	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 6: t-Test for config1 vs. config3 with Equal Variances

	config1	config3
Mean	30.87555555555555	31.397777777777776
Variance	139.15784929757342	162.0026768837803
Observations	30	30
df	58	29
t Stat	-0.16482267270595963	
P(T≤t) two-tail	0.8696566652223652	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

A.1.4 Individual vs. Individual Approach, config1 vs config2

Table 7: F-Test for config1 vs. config2 with $\alpha = 0.05$

	config1	config2
Mean	147.23333333333332	137.5
Variance	333.2367816091954	379.86206896551727
Observations	30	30
df	29	29
F	0.8772573226821593	
P(F≤f) one-tail	0.3633730599975265	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 8: t-Test for config1 vs. config2 with Unequal Variances

	config1	config2
Mean	147.23333333333332	137.5
Variance	333.2367816091954	379.86206896551727
Observations	30	30
df	58	29
t Stat	1.996399056188069	
P(T≤t) two-tail	0.050612077289388574	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

A.1.5 Individual vs. Individual Approach, config2 vs config3

Table 9: F-Test for config2 vs. config3 with $\alpha = 0.05$

	config2	config3
Mean	137.5	147.73333333333332
Variance	379.86206896551727	406.52988505747123
Observations	30	30
df	29	29
F	0.9344013390597743	
P(F≤f) one-tail	0.42814273457822866	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 10: t-Test for config2 vs. config3 with Equal Variances

	config2	config3
Mean	137.5	147.73333333333332
Variance	379.86206896551727	406.52988505747123
Observations	30	30
df	58	29
t Stat	-1.9987488065514472	
P(T≤t) two-tail	0.05032961724630279	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

A.1.6 Individual vs. Individual Approach, config1 vs config3

Table 11: F-Test for config1 vs. config3 with $\alpha = 0.05$

	config1	config3
Mean	147.23333333333332	147.73333333333332
Variance	333.2367816091954	406.52988505747123
Observations	30	30
df	29	29
F	0.8197104170142022	
P(F≤f) one-tail	0.2979622044426937	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 12: t-Test for config1 vs. config3 with Equal Variances

	config1	config3
Mean	147.23333333333332	147.73333333333332
Variance	333.2367816091954	406.52988505747123
Observations	30	30
df	58	29
t Stat	-0.10068928397032623	
P(T≤t) two-tail	0.920144375467848	
t Critical two-tail	2.0017174830120923	

abs(t Stat) < abs(t Critical two-tail) so we accept the null hypothesis – the two samples are NOT statistically different.

A.1.7 "Individual vs. All" Approach vs "Individual vs. Individual" Approach, config1 vs config1

Table 13: F-Test for config1 vs. config1 with $\alpha = 0.05$

	config1	config1
Mean	30.87555555555555	147.23333333333332
Variance	139.15784929757342	333.2367816091954
Observations	30	30
df	29	29
F	0.417594506301442	
P(F≤f) one-tail	0.010841091603319477	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F < F Critical implies unequal variances.

Table 14: t-Test for config1 vs. config1 with Unequal Variances

	config1	config1
Mean	30.87555555555555	147.23333333333332
Variance	139.15784929757342	333.2367816091954
Observations	30	30
df	58	29
t Stat	-29.322673815401515	
P(T≤t) two-tail	5.296450514712298e-33	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of config1 over config1 is -116.35777777777777.

A.2 Cycling Investigation

A.2.1 Individual vs. Individual Approach, config1 vs config4cyc

This section is included for completeness. Given that config4cyc only has 10 runs, this is not a truly statistically-meaningful analysis.

Figure 20: config1 vs. config4cyc – Best values over 30 runs

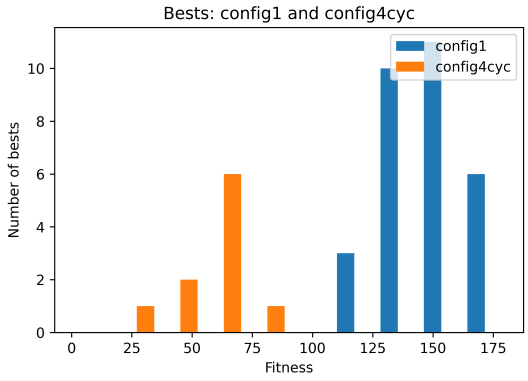


Table 15: F-Test for config1 vs. config4cyc with $\alpha = 0.05$

	config1	config4cyc
Mean	147.23333333333332	59.19
Variance	333.2367816091954	187.51433333333333
Observations	30	30
df	29	29
F	1.7771269837640613	
P(F≤f) one-tail	0.9363271743712435	
F Critical one-tail	0.5373999648406917	

abs(mean 1) > abs(mean 2) and F > F Critical implies unequal variances.

Table 16: t-Test for config1 vs. config4cyc with Unequal Variances

	config1	config4cyc
Mean	147.23333333333332	59.19
Variance	333.2367816091954	187.51433333333333
Observations	30	30
df	58	29
t Stat	16.11226054472109	
P(T≤t) two-tail	3.856883566579471e-13	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of config1 over config4cyc is 88.04333333333332.

A.2.2 Individual vs. Individual Approach, config1 vs config5cyc

This section is included for completeness. Given that config5cyc only has 10 runs, this is not a truly statistically-meaningful analysis.

Figure 21: config1 vs. config5cyc – Best values over 30 runs

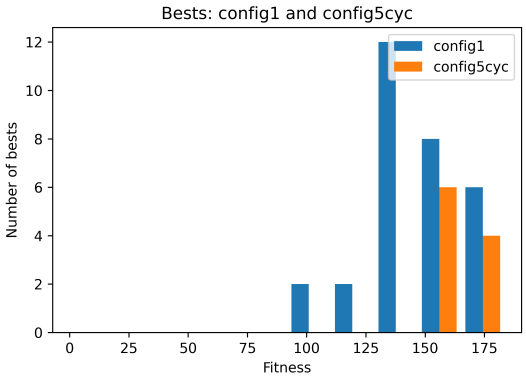


Table 17: F-Test for config1 vs. config5cyc with $\alpha = 0.05$

	config1	config5cyc
Mean	147.23333333333332	165.8
Variance	333.2367816091954	100.67777777777779
Observations	30	30
df	29	29
F	3.3099338201994906	
P(F≤f) one-tail	0.9990568398048119	
F Critical one-tail	0.5373999648406917	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 18: t-Test for config1 vs. config5cyc with Equal Variances

	config1	config5cyc
Mean	147.23333333333332	165.8
Variance	333.2367816091954	100.67777777777779
Observations	30	30
df	58	29
t Stat	-3.048734192916926	
P(T≤t) two-tail	0.0041700613906661	
t Critical two-tail	2.0017174830120923	

abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of config1 over config5cyc is -18.56666666666669.

A.2.3 Individual vs. Individual Approach, config4cyc vs config5cyc

This section is included for completeness. Given that config4cyc and config5cyc only have 10 runs, this is not a truly statistically-meaningful analysis.

Figure 22: config4cyc vs. config5cyc – Best values over 10 runs

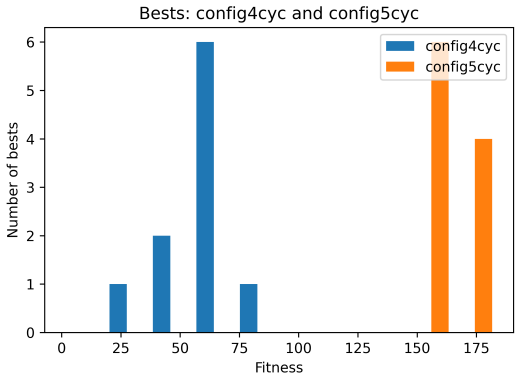


Table 19: F-Test for config4cyc vs. config5cyc with $\alpha = 0.05$

	config4cyc	config5cyc
Mean	59.19	165.8
Variance	187.51433333333333	100.67777777777779
Observations	10	10
df	9	9
F	1.8625195894492879	
P(F≤f) one-tail	0.8160570526741922	
F Critical one-tail	0.31457490615130795	

abs(mean 1) < abs(mean 2) and F > F Critical implies equal variances.

Table 20: t-Test for config4cyc vs. config5cyc with Equal Variances

	config4cyc	config5cyc
Mean	59.19	165.8
Variance	187.51433333333333	100.67777777777779
Observations	10	10
df	18	9
t Stat	-19.85897820406725	
P(T≤t) two-tail	1.0883381286548237e-13	
t Critical two-tail	2.10092204024096	

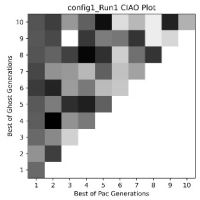
abs(t Stat) > abs(t Critical two-tail) so we reject the null hypothesis – the two samples are statistically different. The average improvement of config4cyc over config5cyc is -106.61000000000001.

B Appendix: CIAO Plots

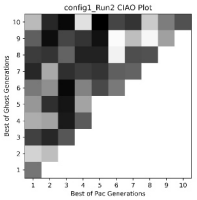
B.1 CIAO Plots for Basic Investigation – Individual vs All Approach

B.1.1 All CIAO plots for config1

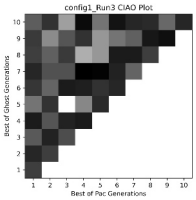
All CIAO plots for the 30 runs of config1 are on the next page.



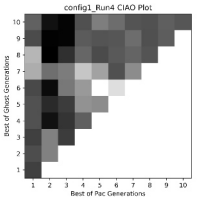
config1_Run1_CIAO_Plot



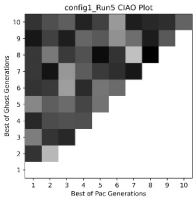
config1_Run2_CIAO_Plot



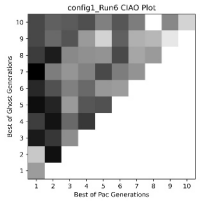
config1_Run3_CIAO_Plot



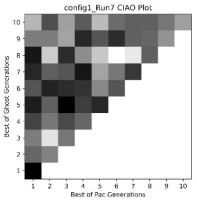
config1_Run4_CIAO_Plot



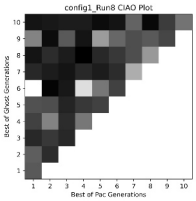
config1_Run5_CIAO_Plot



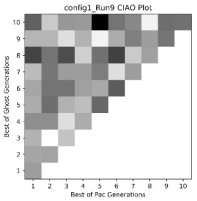
config1_Run6_CIAO_Plot



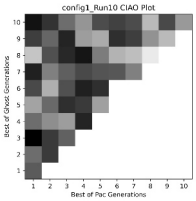
config1_Run7_CIAO_Plot



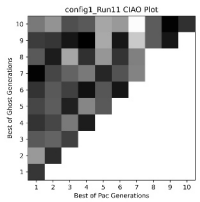
config1_Run8_CIAO_Plot



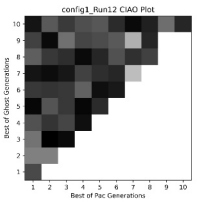
config1_Run9_CIAO_Plot



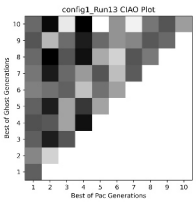
config1_Run10_CIAO_Plot



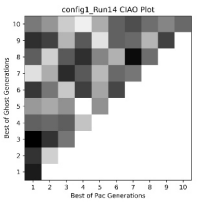
config1_Run11_CIAO_Plot



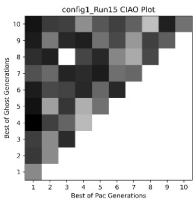
config1_Run12_CIAO_Plot



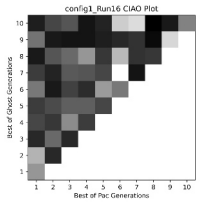
config1_Run13_CIAO_Plot



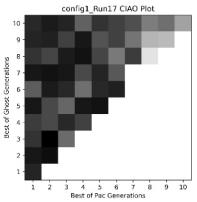
config1_Run14_CIAO_Plot



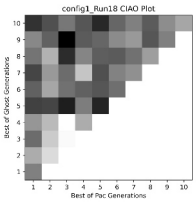
config1_Run15_CIAO_Plot



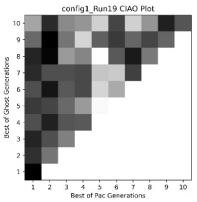
config1_Run16_CIAO_Plot



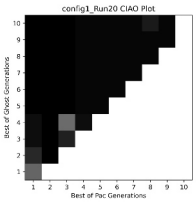
config1_Run17_CIAO_Plot



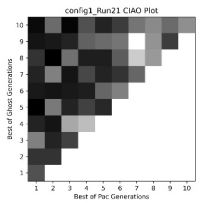
config1_Run18_CIAO_Plot



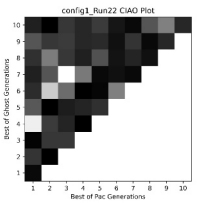
config1_Run19_CIAO_Plot



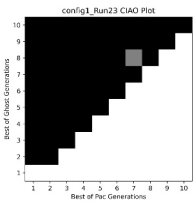
config1_Run20_CIAO_Plot



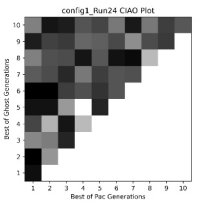
config1_Run21_CIAO_Plot



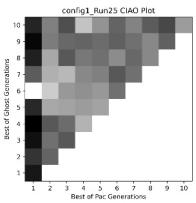
config1_Run22_CIAO_Plot



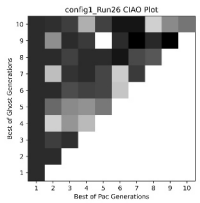
config1_Run23_CIAO_Plot



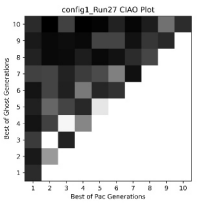
config1_Run24_CIAO_Plot



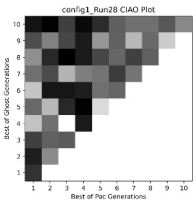
config1_Run25_CIAO_Plot



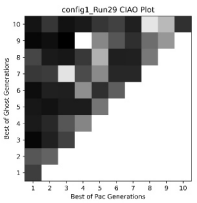
config1_Run26_CIAO_Plot



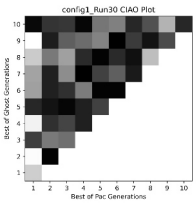
config1_Run27_CIAO_Plot



config1_Run28_CIAO_Plot



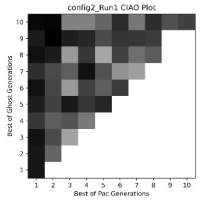
config1_Run29_CIAO_Plot



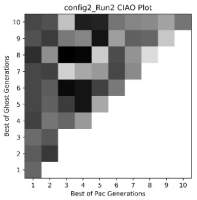
config1_Run30_CIAO_Plot

B.1.2 All CIAO plots for config2

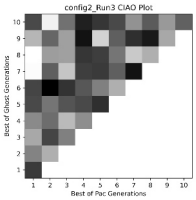
All CIAO plots for the 30 runs of config2 are on the next page.



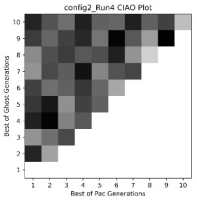
config2_Run1_CIAO_Plot



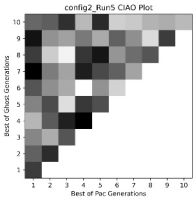
config2_Run2_CIAO_Plot



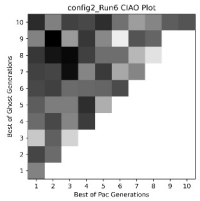
config2_Run3_CIAO_Plot



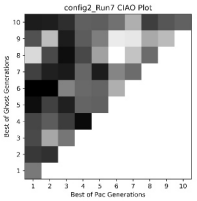
config2_Run4_CIAO_Plot



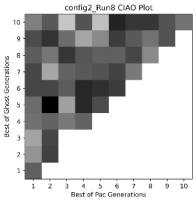
config2_Run5_CIAO_Plot



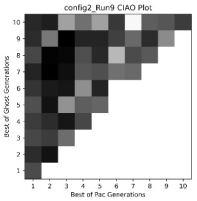
config2_Run6_CIAO_Plot



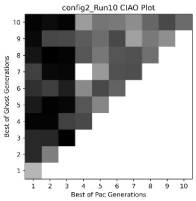
config2_Run7_CIAO_Plot



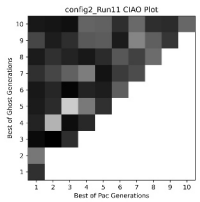
config2_Run8_CIAO_Plot



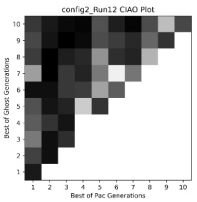
config2_Run9_CIAO_Plot



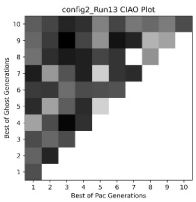
config2_Run10_CIAO_Plot



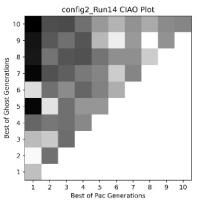
config2_Run11_CIAO_Plot



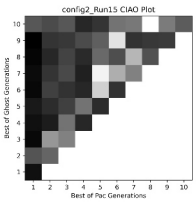
config2_Run12_CIAO_Plot



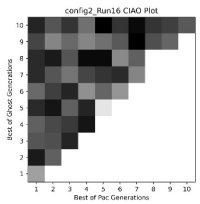
config2_Run13_CIAO_Plot



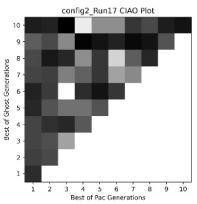
config2_Run14_CIAO_Plot



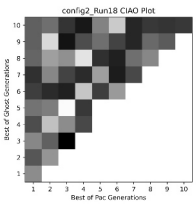
config2_Run15_CIAO_Plot



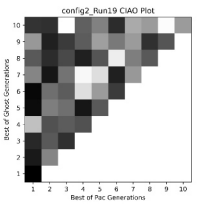
config2_Run16_CIAO_Plot



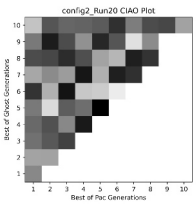
config2_Run17_CIAO_Plot



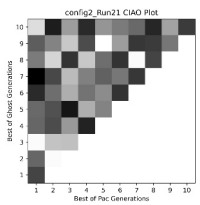
config2_Run18_CIAO_Plot



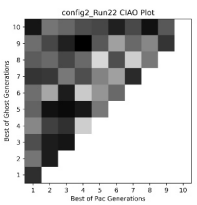
config2_Run19_CIAO_Plot



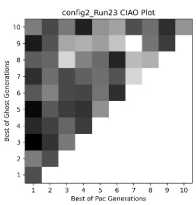
config2_Run20_CIAO_Plot



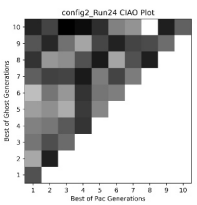
config2_Run21_CIAO_Plot



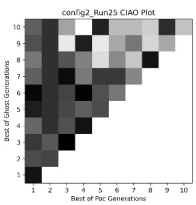
config2_Run22_CIAO_Plot



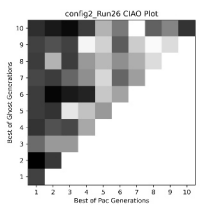
config2_Run23_CIAO_Plot



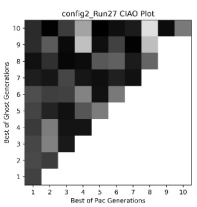
config2_Run24_CIAO_Plot



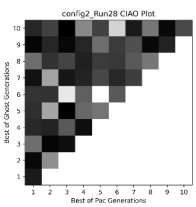
config2_Run25_CIAO_Plot



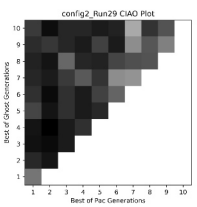
config2_Run26_CIAO_Plot



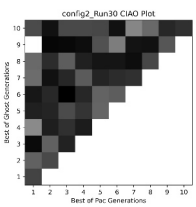
config2_Run27_CIAO_Plot



config2_Run28_CIAO_Plot



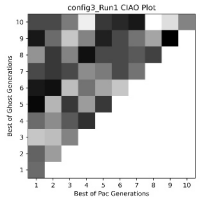
config2_Run29_CIAO_Plot



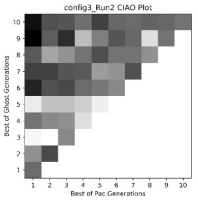
config2_Run30_CIAO_Plot

B.1.3 All CIAO plots for config3

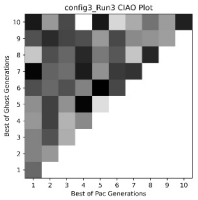
All CIAO plots for the 30 runs of config3 are on the next page.



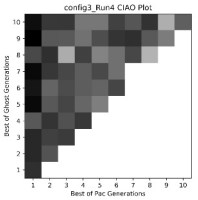
config3_Run1_CIAO_Plot



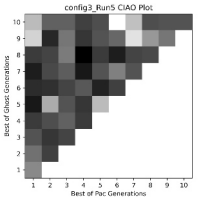
config3_Run2_CIAO_Plot



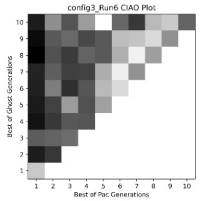
config3_Run3_CIAO_Plot



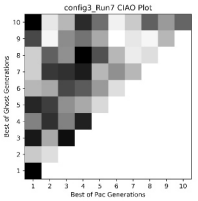
config3_Run4_CIAO_Plot



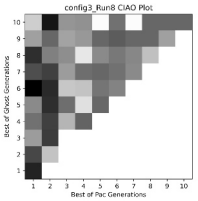
config3_Run5_CIAO_Plot



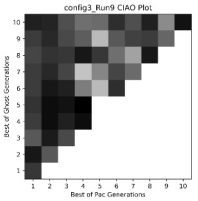
config3_Run6_CIAO_Plot



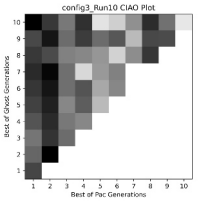
config3_Run7_CIAO_Plot



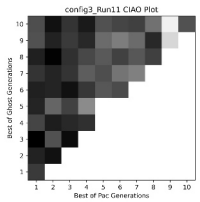
config3_Run8_CIAO_Plot



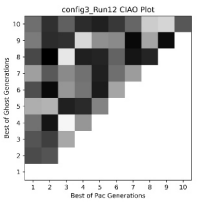
config3_Run9_CIAO_Plot



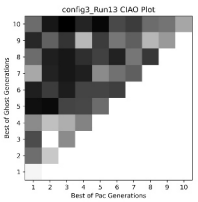
config3_Run10_CIAO_Plot



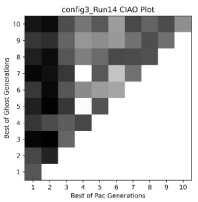
config3_Run11_CIAO_Plot



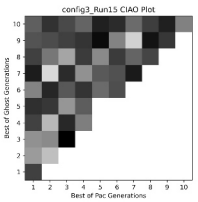
config3_Run12_CIAO_Plot



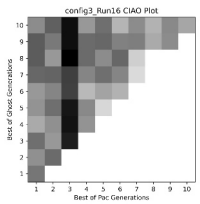
config3_Run13_CIAO_Plot



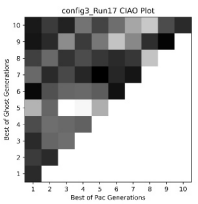
config3_Run14_CIAO_Plot



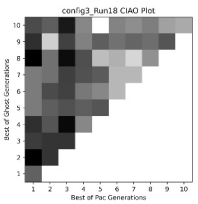
config3_Run15_CIAO_Plot



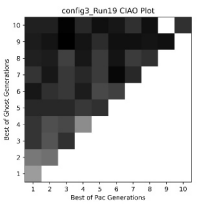
config3_Run16_CIAO_Plot



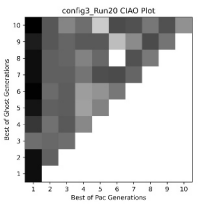
config3_Run17_CIAO_Plot



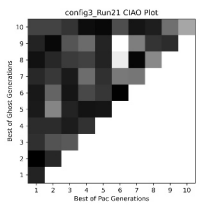
config3_Run18_CIAO_Plot



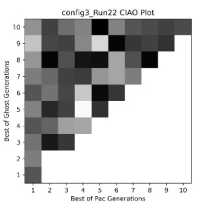
config3_Run19_CIAO_Plot



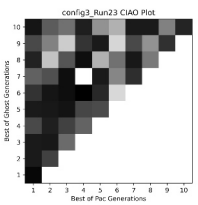
config3_Run20_CIAO_Plot



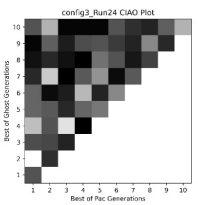
config3_Run21_CIAO_Plot



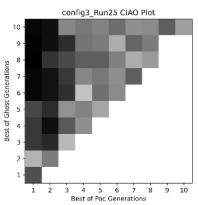
config3_Run22_CIAO_Plot



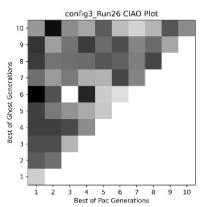
config3_Run23_CIAO_Plot



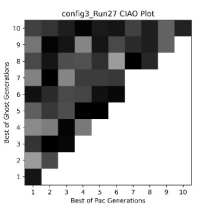
config3_Run24_CIAO_Plot



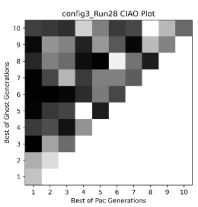
config3_Run25_CIAO_Plot



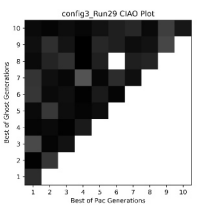
config3_Run26_CIAO_Plot



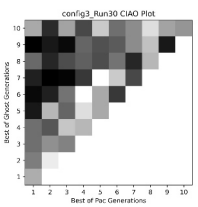
config3_Run27_CIAO_Plot



config3_Run28_CIAO_Plot



config3_Run29_CIAO_Plot

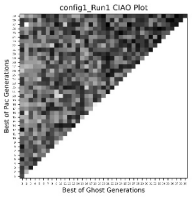


config3_Run30_CIAO_Plot

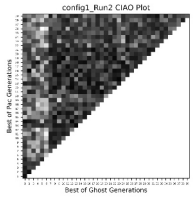
B.2 CIAO Plots for Basic Investigation – Individual vs Individual Approach

B.2.1 All CIAO plots for config1

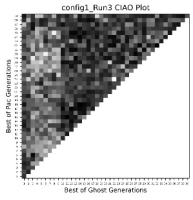
All CIAO plots for the 30 runs of config1 are on the next page.



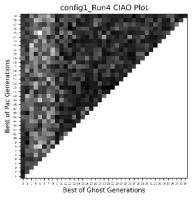
config1_Run1_CIAO_Plot



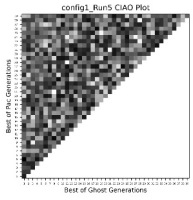
config1_Run2_CIAO_Plot



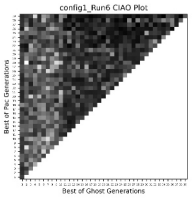
config1_Run3_CIAO_Plot



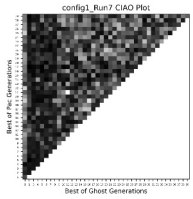
config1_Run4_CIAO_Plot



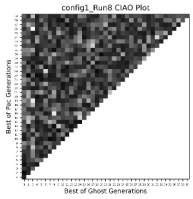
config1_Run5_CIAO_Plot



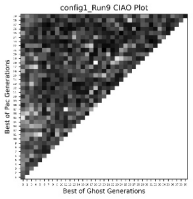
config1_Run6_CIAO_Plot



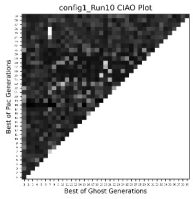
config1_Run7_CIAO_Plot



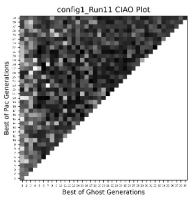
config1_Run8_CIAO_Plot



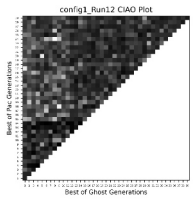
config1_Run9_CIAO_Plot



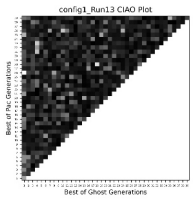
config1_Run10_CIAO_Plot



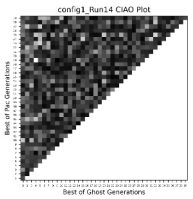
config1_Run11_CIAO_Plot



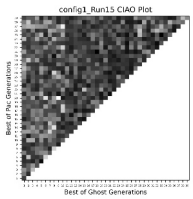
config1_Run12_CIAO_Plot



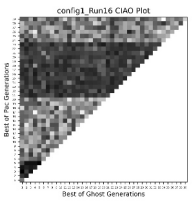
config1_Run13_CIAO_Plot



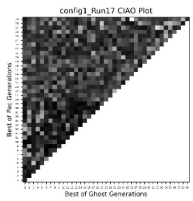
config1_Run14_CIAO_Plot



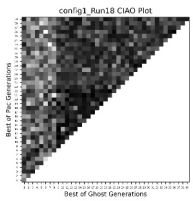
config1_Run15_CIAO_Plot



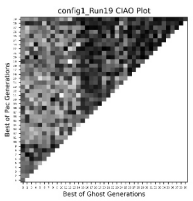
config1_Run16_CIAO_Plot



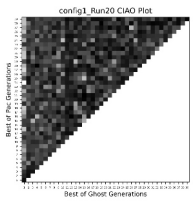
config1_Run17_CIAO_Plot



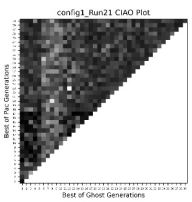
config1_Run18_CIAO_Plot



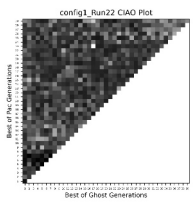
config1_Run19_CIAO_Plot



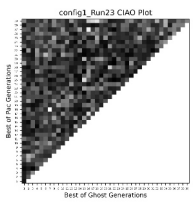
config1_Run20_CIAO_Plot



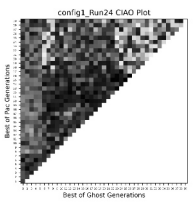
config1_Run21_CIAO_Plot



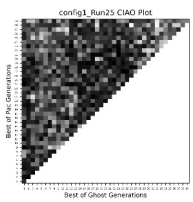
config1_Run22_CIAO_Plot



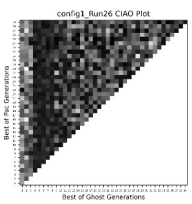
config1_Run23_CIAO_Plot



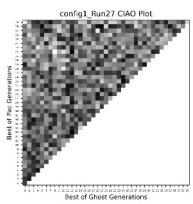
config1_Run24_CIAO_Plot



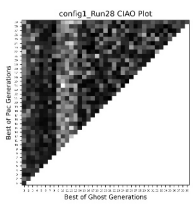
config1_Run25_CIAO_Plot



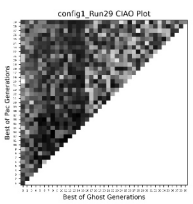
config1_Run26_CIAO_Plot



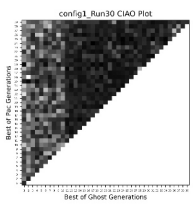
config1_Run27_CIAO_Plot



config1_Run28_CIAO_Plot



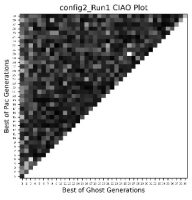
config1_Run29_CIAO_Plot



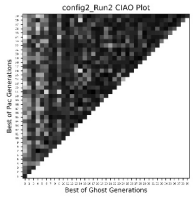
config1_Run30_CIAO_Plot

B.2.2 All CIAO plots for config2

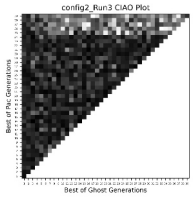
All CIAO plots for the 30 runs of config2 are on the next page.



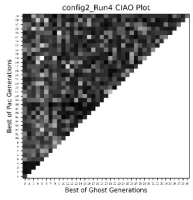
config2_Run1_CIAO_Plot



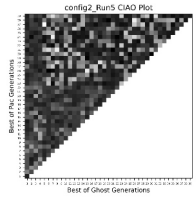
config2_Run2_CIAO_Plot



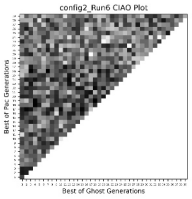
config2_Run3_CIAO_Plot



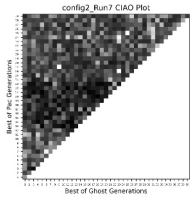
config2_Run4_CIAO_Plot



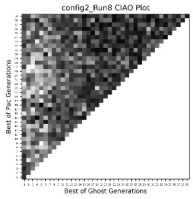
config2_Run5_CIAO_Plot



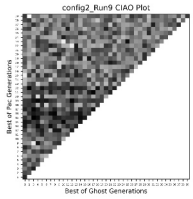
config2_Run6_CIAO_Plot



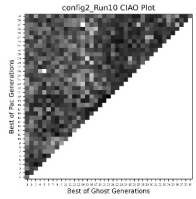
config2_Run7_CIAO_Plot



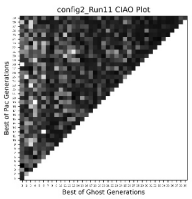
config2_Run8_CIAO_Plot



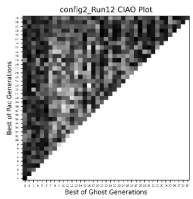
config2_Run9_CIAO_Plot



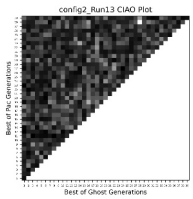
config2_Run10_CIAO_Plot



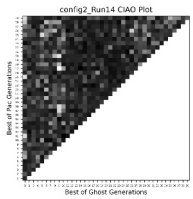
config2_Run11_CIAO_Plot



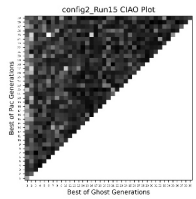
config2_Run12_CIAO_Plot



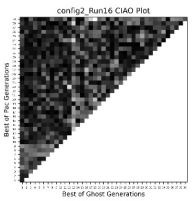
config2_Run13_CIAO_Plot



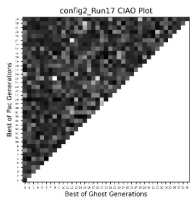
config2_Run14_CIAO_Plot



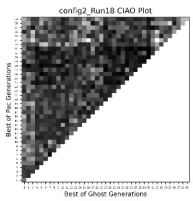
config2_Run15_CIAO_Plot



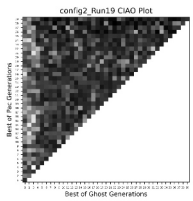
config2_Run16_CIAO_Plot



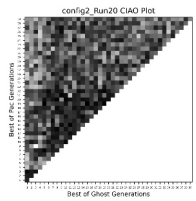
config2_Run17_CIAO_Plot



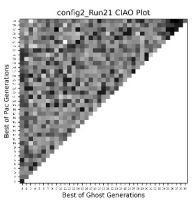
config2_Run18_CIAO_Plot



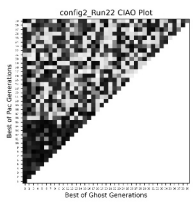
config2_Run19_CIAO_Plot



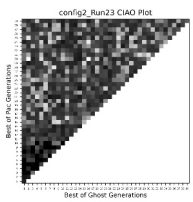
config2_Run20_CIAO_Plot



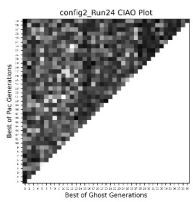
config2_Run21_CIAO_Plot



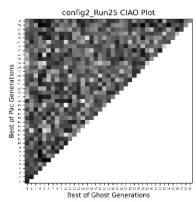
config2_Run22_CIAO_Plot



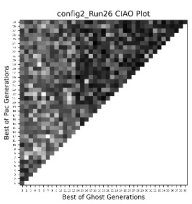
config2_Run23_CIAO_Plot



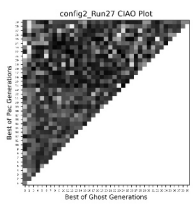
config2_Run24_CIAO_Plot



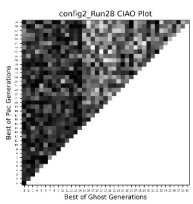
config2_Run25_CIAO_Plot



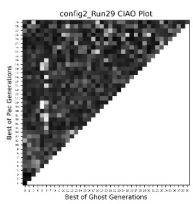
config2_Run26_CIAO_Plot



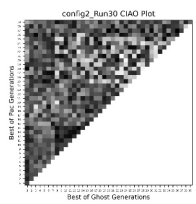
config2_Run27_CIAO_Plot



config2_Run28_CIAO_Plot



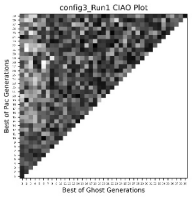
config2_Run29_CIAO_Plot



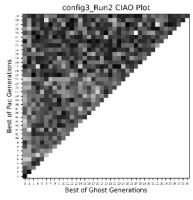
config2_Run30_CIAO_Plot

B.2.3 All CIAO plots for config3

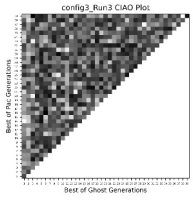
All CIAO plots for the 30 runs of config3 are on the next page.



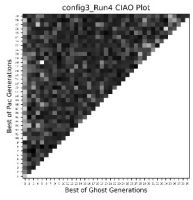
config3_Run1_CIAO_Plot



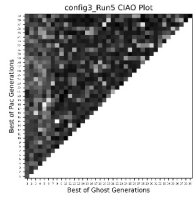
config3_Run2_CIAO_Plot



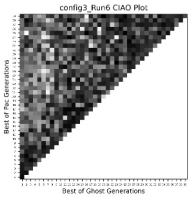
config3_Run3_CIAO_Plot



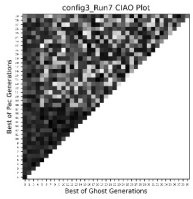
config3_Run4_CIAO_Plot



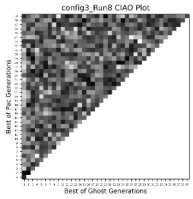
config3_Run5_CIAO_Plot



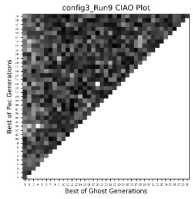
config3_Run6_CIAO_Plot



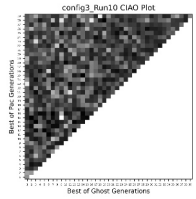
config3_Run7_CIAO_Plot



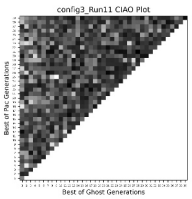
config3_Run8_CIAO_Plot



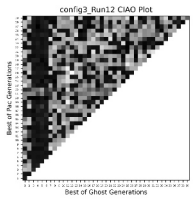
config3_Run9_CIAO_Plot



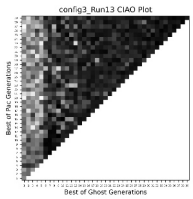
config3_Run10_CIAO_Plot



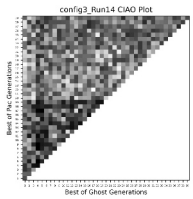
config3_Run11_CIAO_Plot



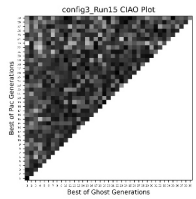
config3_Run12_CIAO_Plot



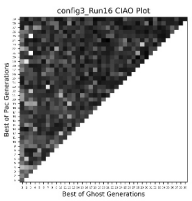
config3_Run13_CIAO_Plot



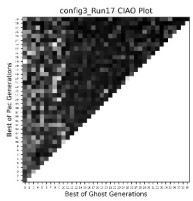
config3_Run14_CIAO_Plot



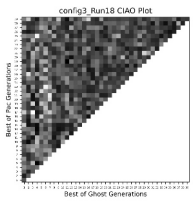
config3_Run15_CIAO_Plot



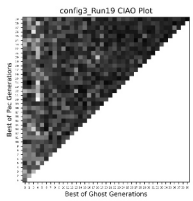
config3_Run16_CIAO_Plot



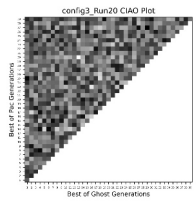
config3_Run17_CIAO_Plot



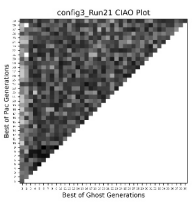
config3_Run18_CIAO_Plot



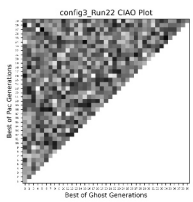
config3_Run19_CIAO_Plot



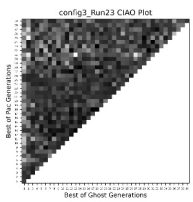
config3_Run20_CIAO_Plot



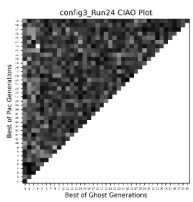
config3_Run21_CIAO_Plot



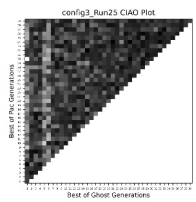
config3_Run22_CIAO_Plot



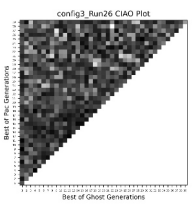
config3_Run23_CIAO_Plot



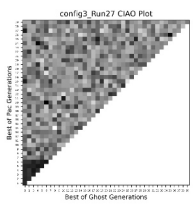
config3_Run24_CIAO_Plot



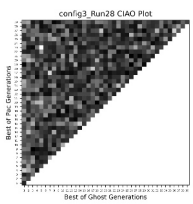
config3_Run25_CIAO_Plot



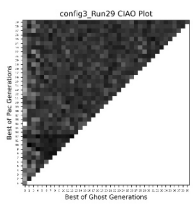
config3_Run26_CIAO_Plot



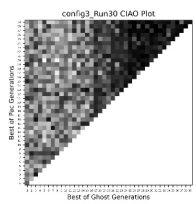
config3_Run27_CIAO_Plot



config3_Run28_CIAO_Plot



config3_Run29_CIAO_Plot

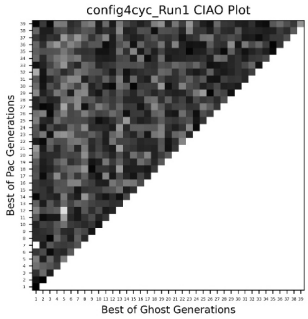


config3_Run30_CIAO_Plot

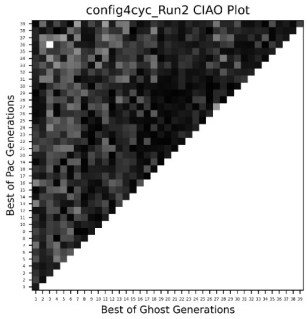
B.3 CIAO Plots for Cycling Investigation

B.3.1 All CIAO plots for config4cyc

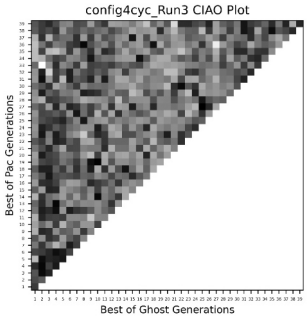
All CIAO plots for the 30 runs of config4cyc are on the next page.



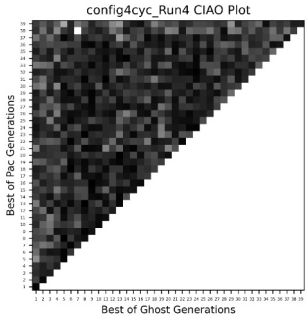
config4cyc_Run1_CIAO_Plot



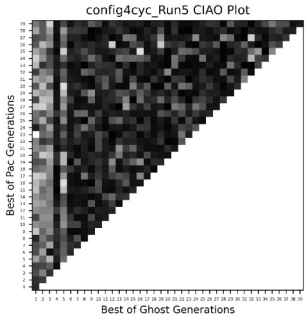
config4cyc_Run2_CIAO_Plot



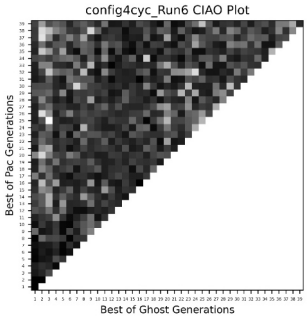
config4cyc_Run3_CIAO_Plot



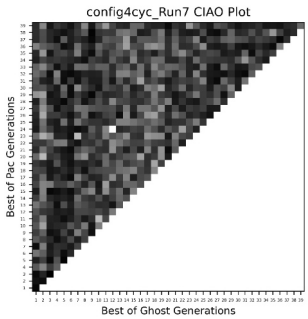
config4cyc_Run4_CIAO_Plot



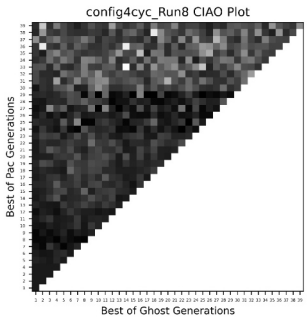
config4cyc_Run5_CIAO_Plot



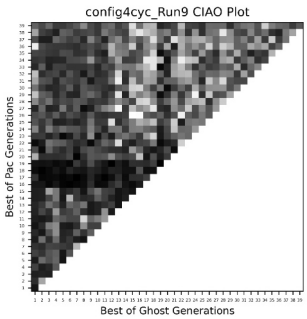
config4cyc_Run6_CIAO_Plot



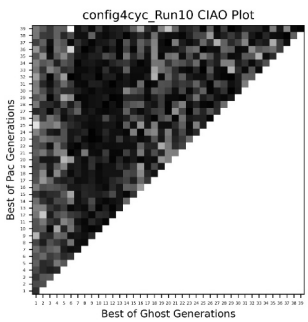
config4cyc_Run7_CIAO_Plot



config4cyc_Run8_CIAO_Plot



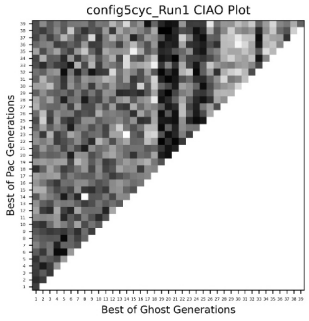
config4cyc_Run9_CIAO_Plot



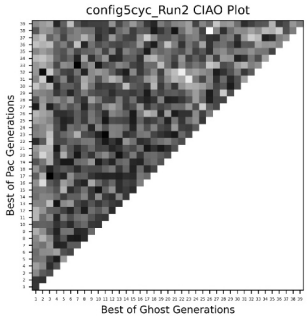
config4cyc_Run10_CIAO_Plot

B.3.2 All CIAO plots for config5cyc

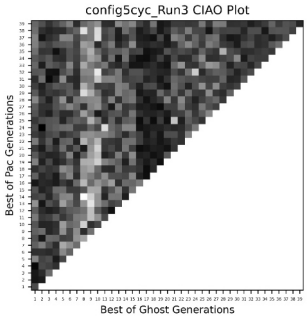
All CIAO plots for the 30 runs of config5cyc are on the next page.



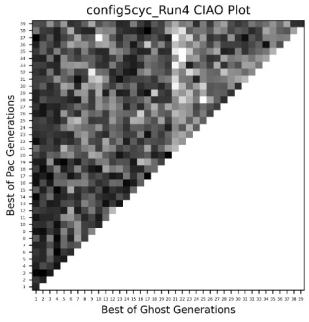
config5cyc_Run1_CIAO_Plot



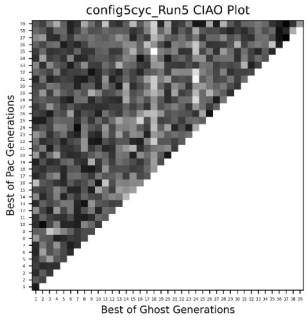
config5cyc_Run2_CIAO_Plot



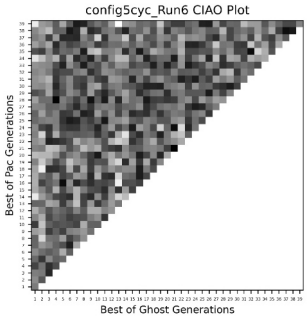
config5cyc_Run3_CIAO_Plot



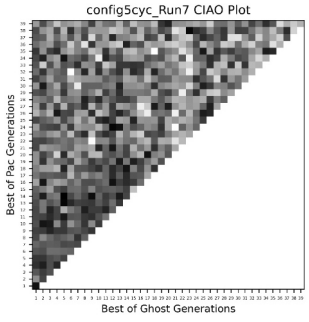
config5cyc_Run4_CIAO_Plot



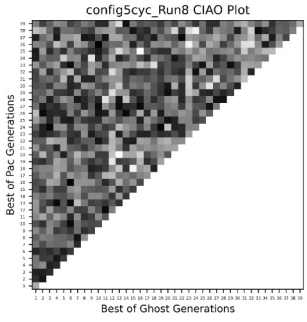
config5cyc_Run5_CIAO_Plot



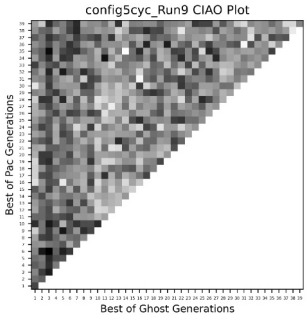
config5cyc_Run6_CIAO_Plot



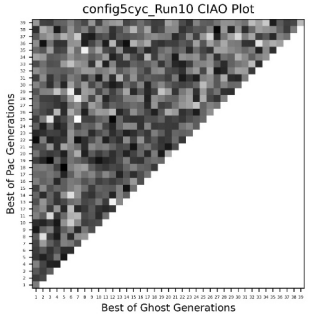
config5cyc_Run7_CIAO_Plot



config5cyc_Run8_CIAO_Plot



config5cyc_Run9_CIAO_Plot



config5cyc_Run10_CIAO_Plot