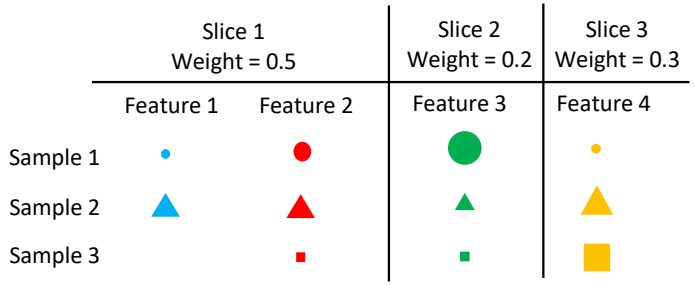
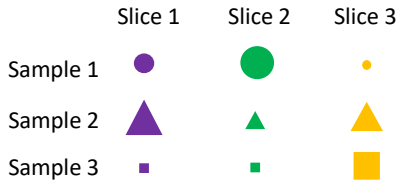
User Defined Model For Feature Data



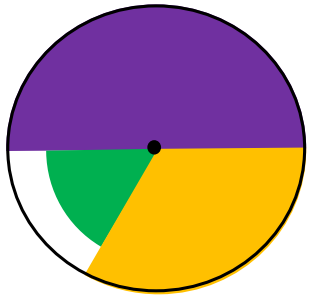
Aggregate Features Into Slices



Rescale Slice Scores 0-1

Calculate Sample Scores

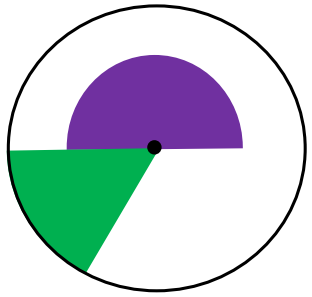
Sample 2
Score = 0.95
Rank = 1/3



Sample 2:

- Highest overall outcome
- Highest Slice 1 association
- Moderate Slice 2 association
- Highest Slice 3 association
- Highest prioritized sample driven by all features

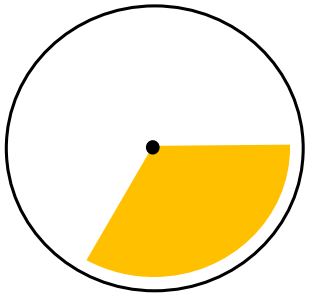
Sample 1
Score = 0.45
Rank = 2/3



Sample 1:

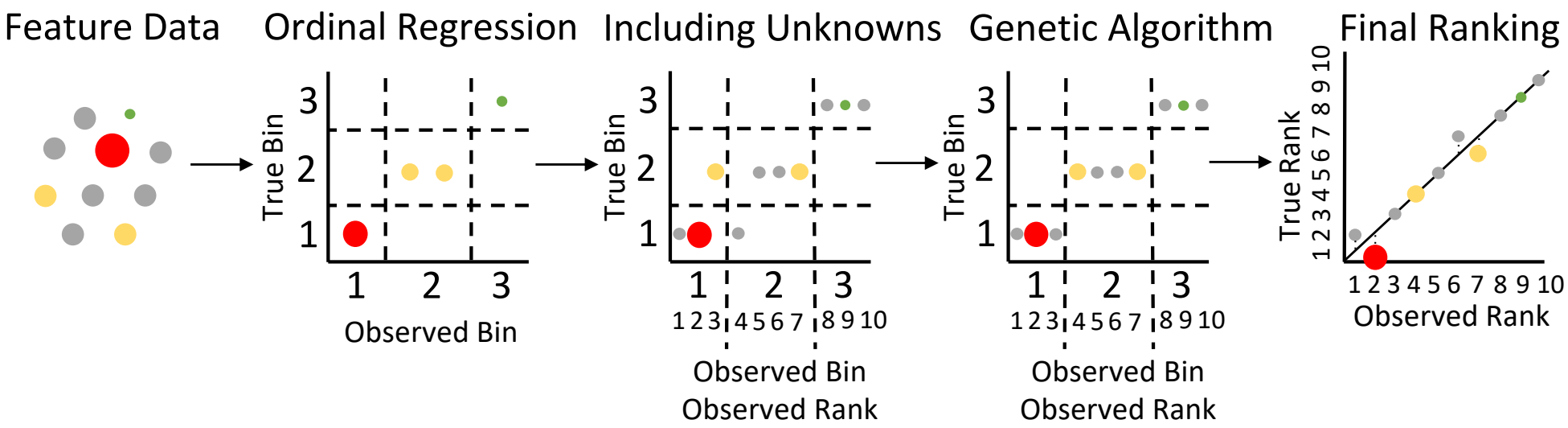
- Moderate overall outcome
- Moderate Slice 1 association
- Highest Slice 2 association
- Lowest Slice 3 association
- Moderately prioritized sample driven largely by Slice 2 features

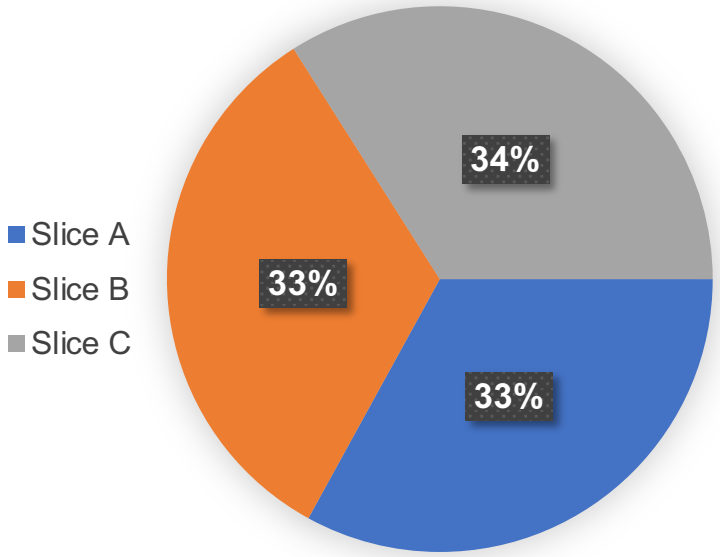
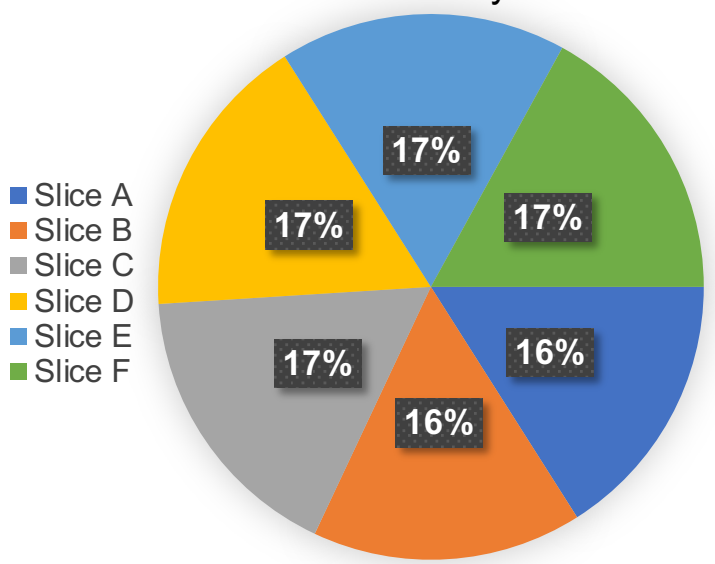
Sample 3
Score = 0.25
Rank = 3/3



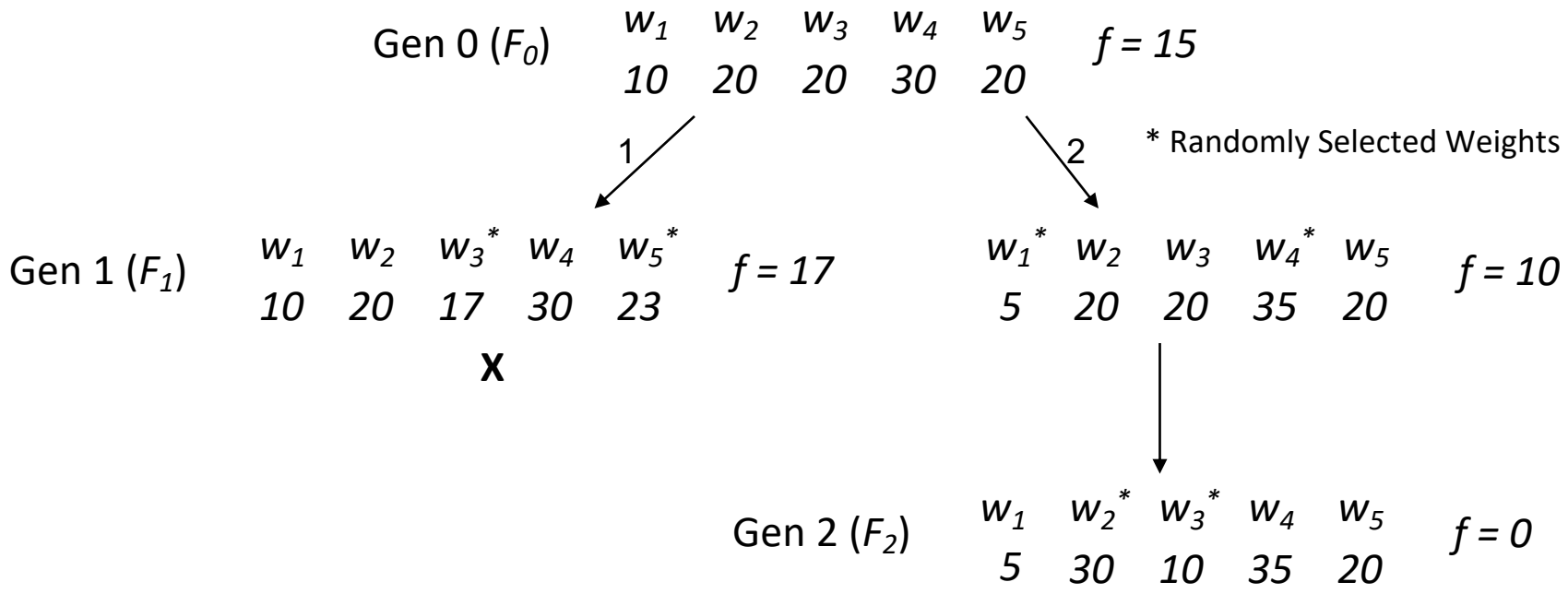
Sample 3:

- Lowest overall outcome
- Lowest Slice 1 association
- Lowest Slice 2 association
- High Slice 3 association
- Lowly prioritized sample, special concerns for Slice 3 features

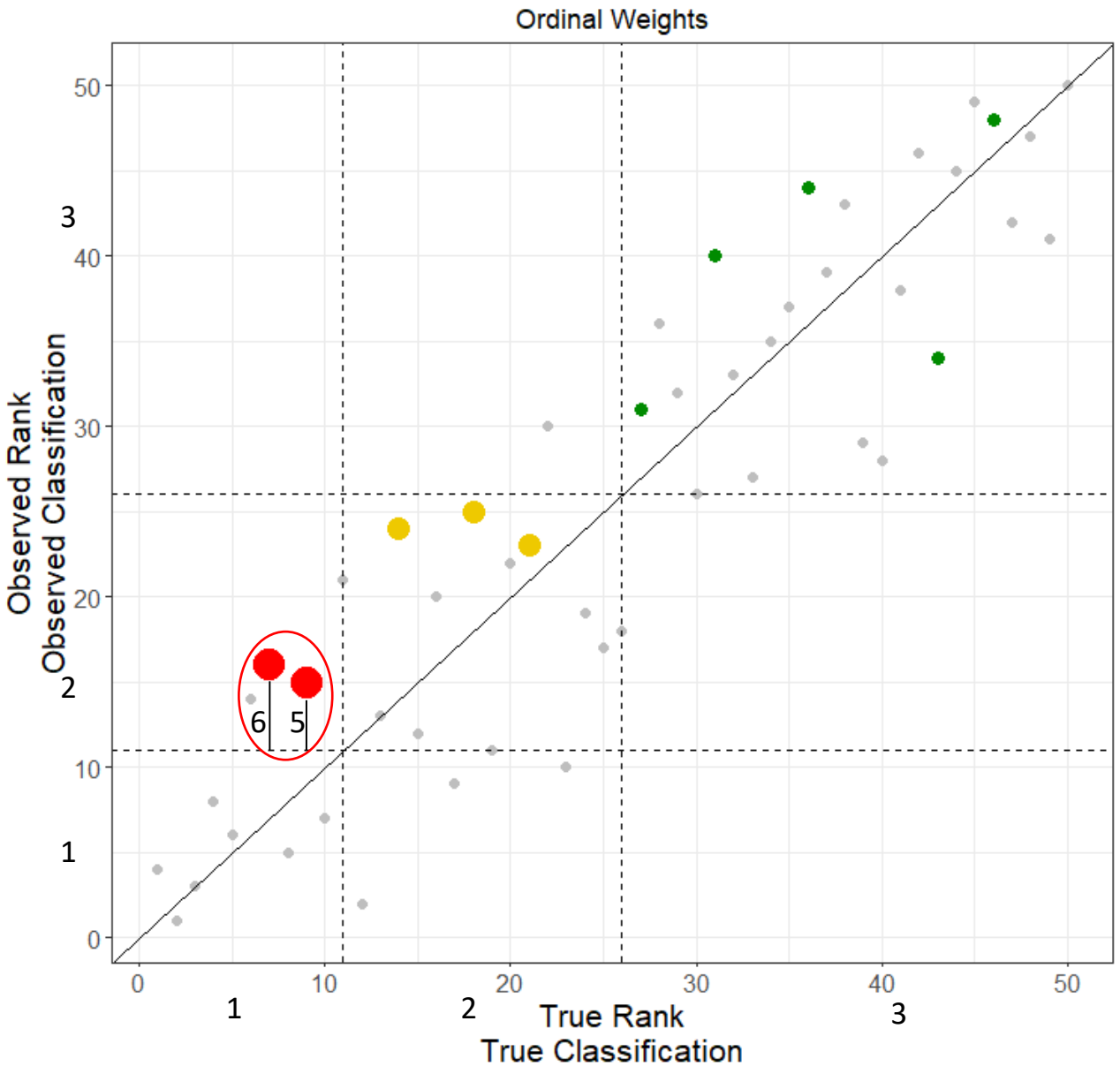

Feature Data
Ordinal Regression
Including Unknowns
Genetic Algorithm
Final Ranking
True Bin
Observed Bin
Observed Rank
True Rank

Combination 1: 3 Slices, 3 Knowns/Slice, 500 Samples, 2 Bins	Combination 2: 6 Slices, 3 Knowns/Slice, 1000 Samples, 3 Bins
ToxPi Profile Layout  Data Layout Known Samples Unknown Samples Total Samples 3 x 3 + x 491 = 500	ToxPi Profile Layout  Data Layout Known Samples Unknown Samples Total Samples 6 x 3 + x 982 = 1000

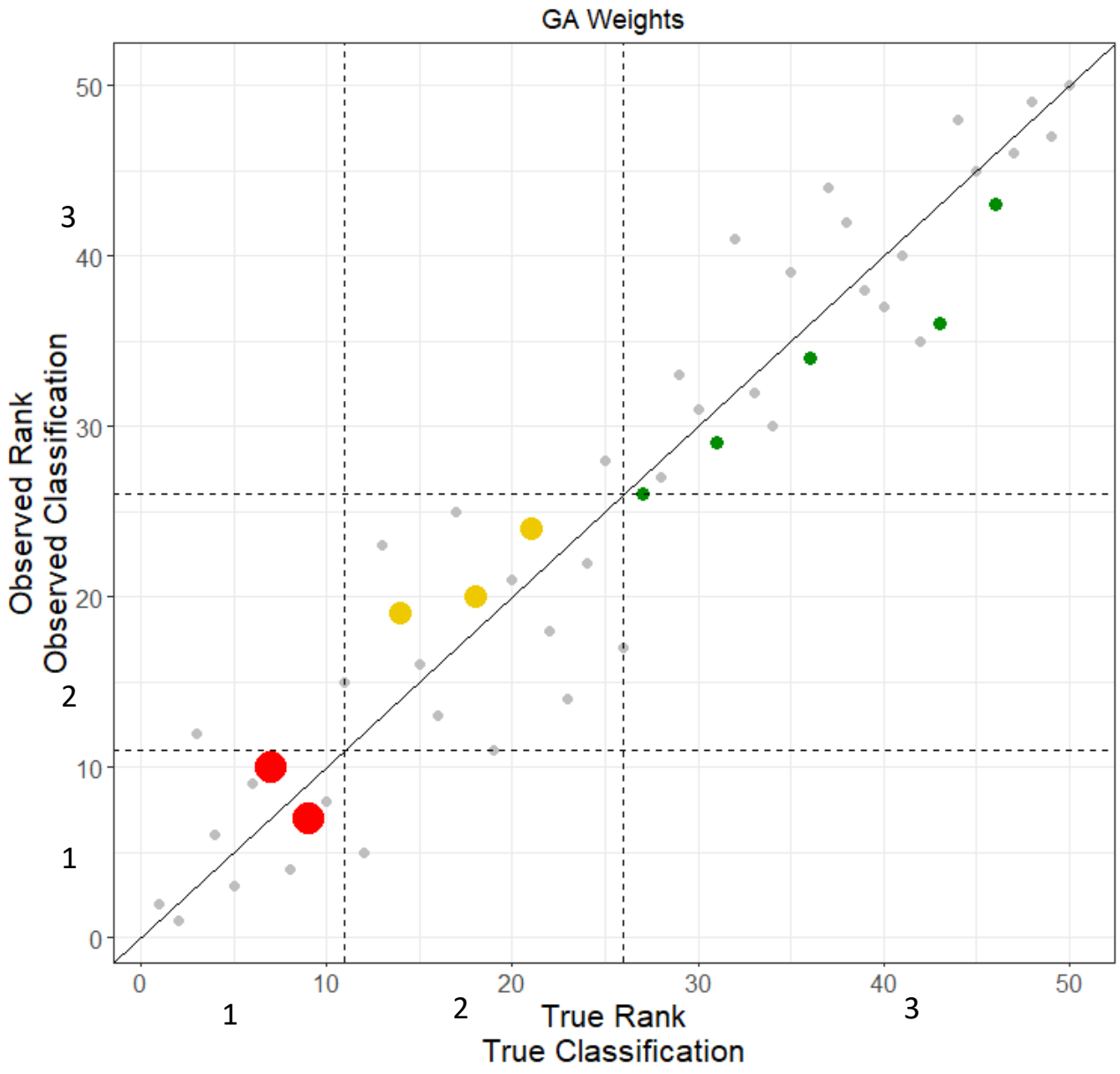
Slices:	Knowns/Slice:	Total Samples:	Bins:	Underlying Data Distribution:
{3,6,9,12,15}	{3,6,9,12,15}	{500,1000,5000,10000}	{2,3,4}	{Normal, Gamma, Beta, Mixed}



Final w = {5, 30, 10, 35, 20}

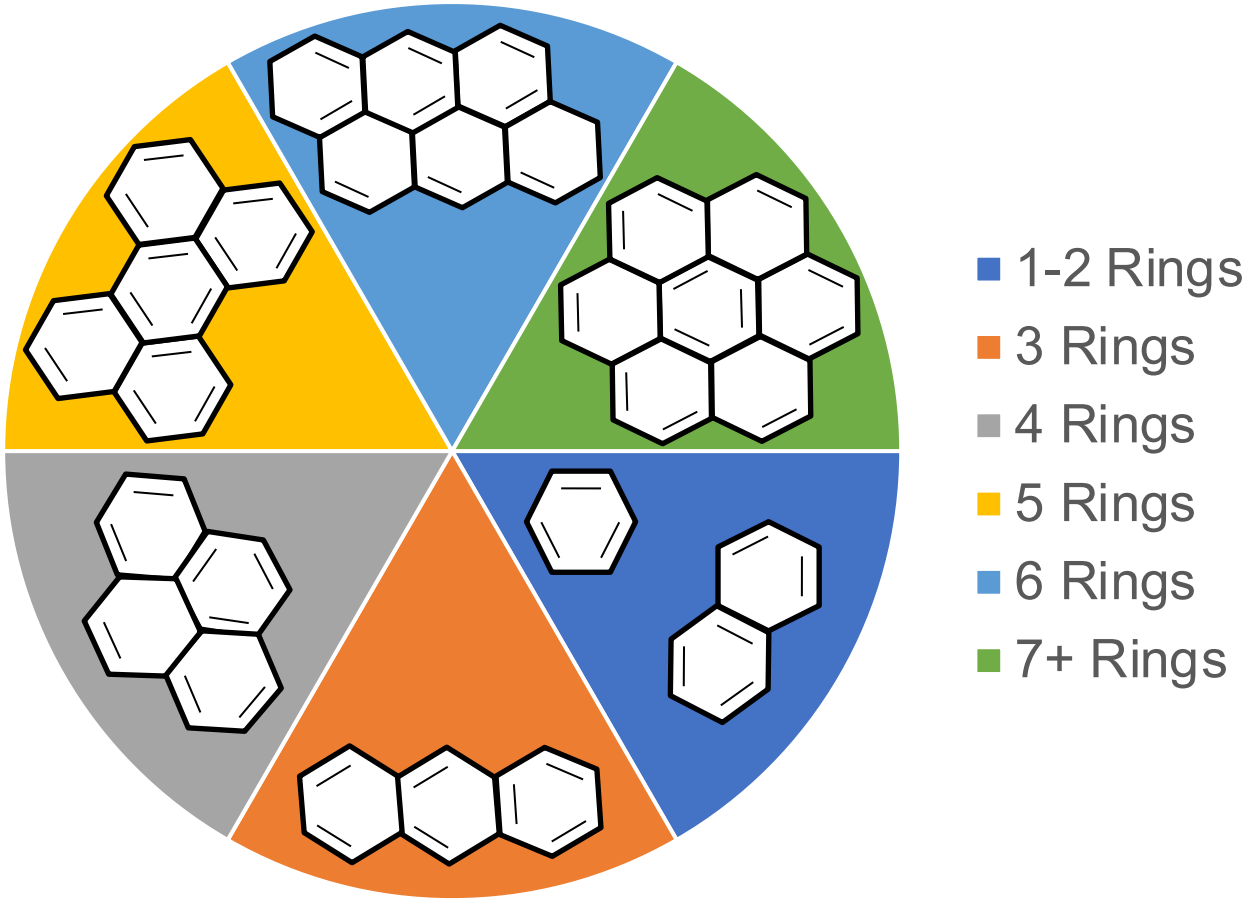


$$f = (2-1)*6 + (2-1)*5 = 11$$

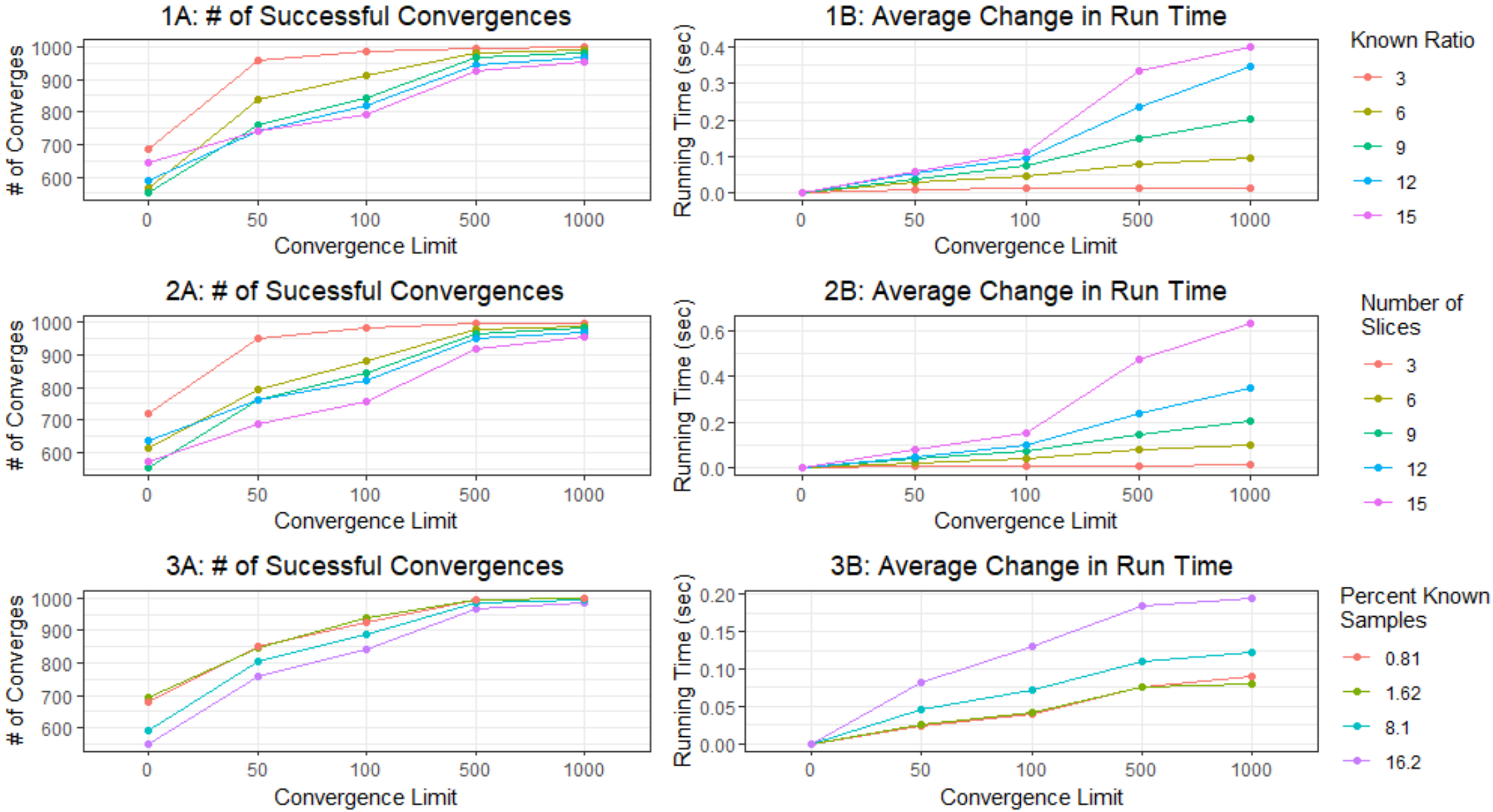


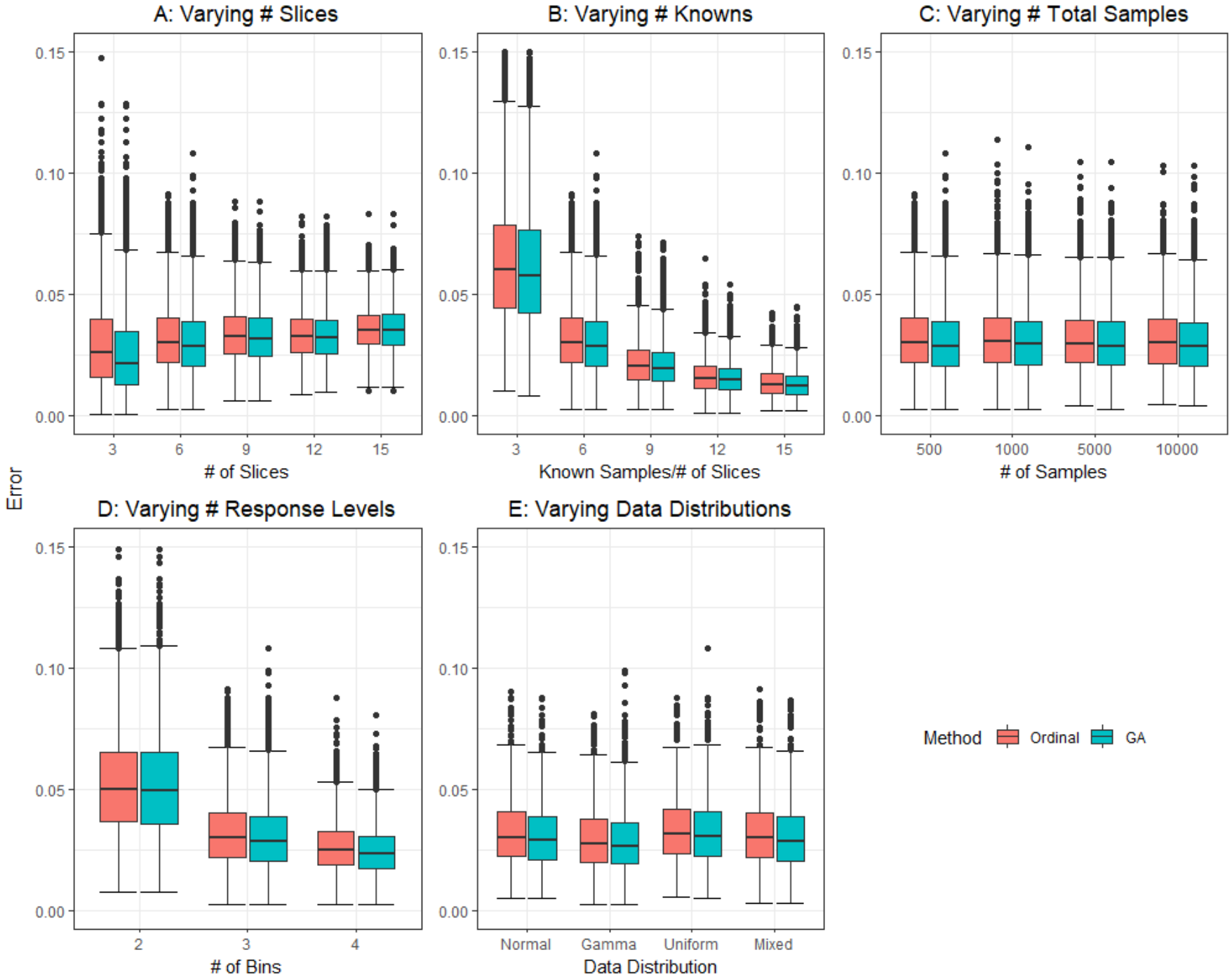
$$f = 0$$

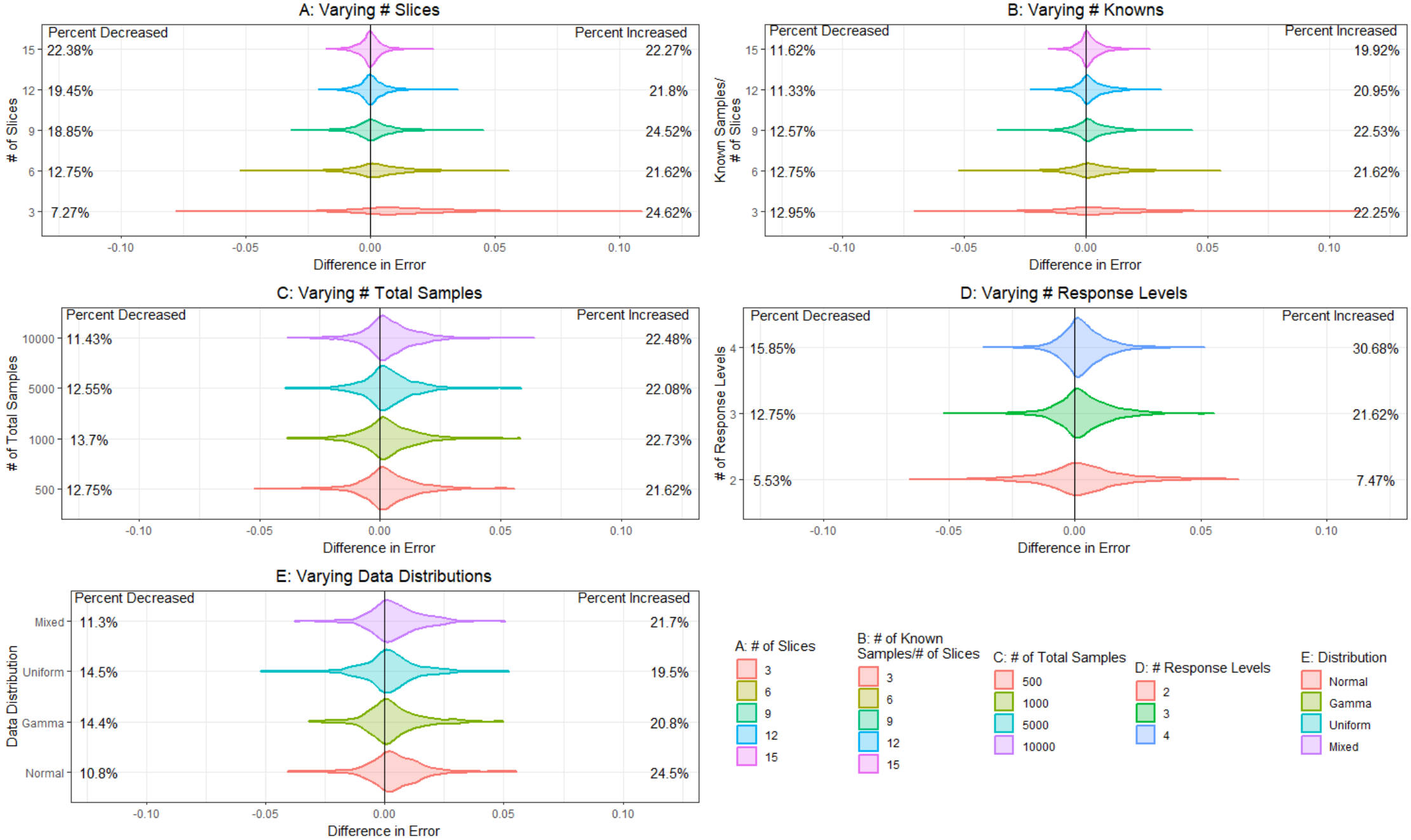
PAC Data ToxPi Profile

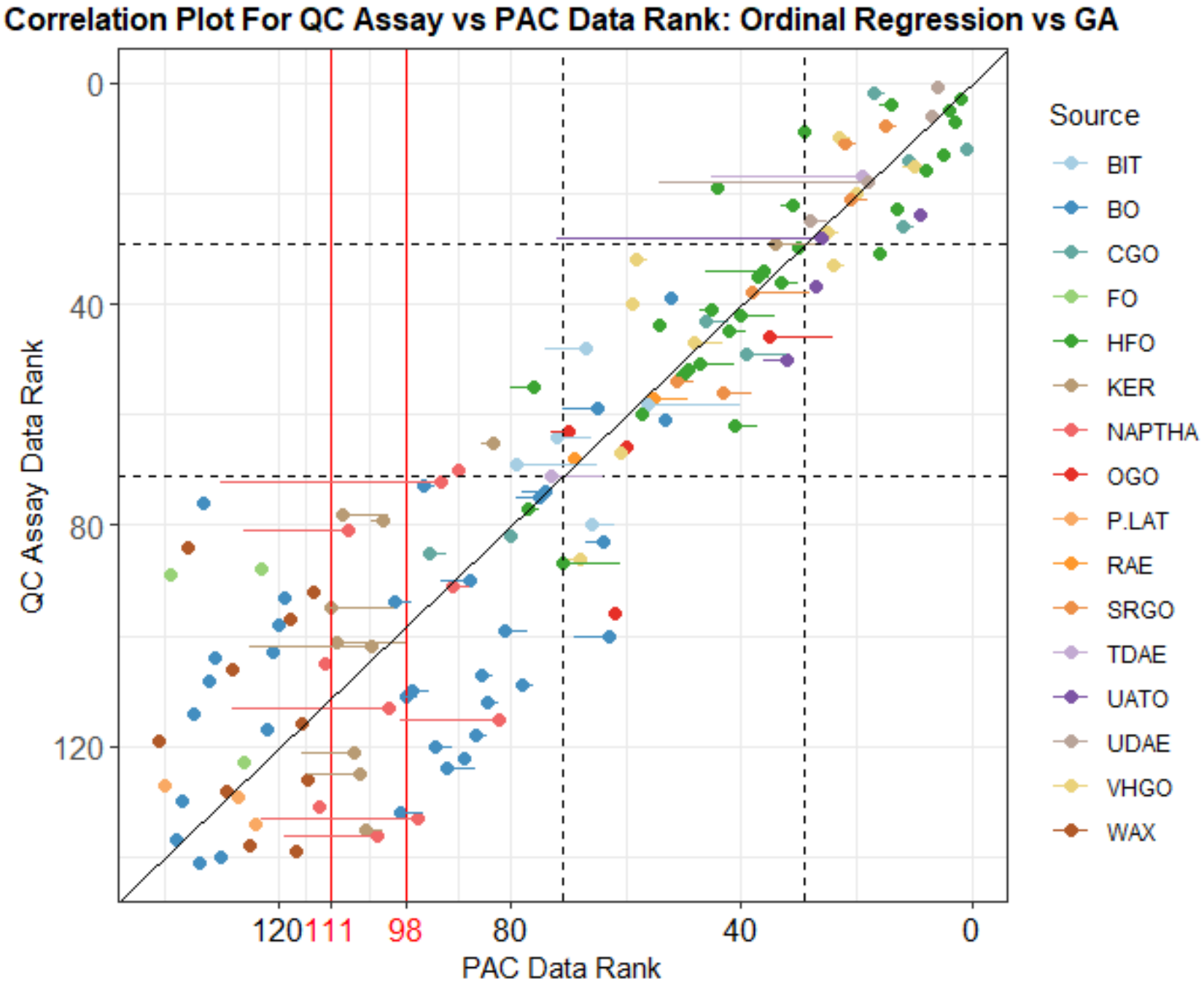


- 1-2 Rings
- 3 Rings
- 4 Rings
- 5 Rings
- 6 Rings
- 7+ Rings


1A: # of Successful Convergences
# of Converges
1000
900
800
700
600
0
50
100
500
1000
Convergence Limit
1B: Average Change in Run Time
Running Time (sec)
0.4
0.3
0.2
0.1
0.0
Convergence Limit
Known Ratio
3
6
9
12
15
2A: # of Sucessful Convergences
2B: Average Change in Run Time
0.6
0.4
0.2
0.0
Number of Slices
3A: # of Sucessful Convergences
3B: Average Change in Run Time
0.20
0.15
0.10
0.05
0.00
Percent Known Samples
0.81
1.62
8.1
16.2


A: Varying # Slices
B: Varying # Knowns
C: Varying # Total Samples
D: Varying # Response Levels
E: Varying Data Distributions
Error
0.15
0.10
0.05
0.00
3
6
9
12
15
# of Slices
Known Samples/# of Slices
500
1000
5000
10000
# of Samples
2
3
4
# of Bins
Normal
Gamma
Uniform
Mixed
Data Distribution
Method
Ordinal
GA


A: Varying # Slices
Percent Decreased
Percent Increased
# of Slices
15
12
9
6
3
22.38%
19.45%
18.85%
12.75%
7.27%
22.27%
21.8%
24.52%
21.62%
24.62%
-0.10
-0.05
0.00
0.05
0.10
Difference in Error
B: Varying # Knowns
Known Samples/ # of Slices
11.62%
11.33%
12.57%
12.75%
12.95%
19.92%
20.95%
22.53%
21.62%
22.25%
C: Varying # Total Samples
# of Total Samples
10000
5000
1000
500
11.43%
12.55%
13.7%
12.75%
22.48%
22.08%
22.73%
21.62%
D: Varying # Response Levels
# of Response Levels
4
3
2
15.85%
12.75%
5.53%
30.68%
21.62%
7.47%
E: Varying Data Distributions
Data Distribution
Mixed
Uniform
Gamma
Normal
11.3%
14.5%
14.4%
10.8%
21.7%
19.5%
20.8%
24.5%
A: # of Slices
3
6
9
12
15
B: # of Known Samples/# of Slices
C: # of Total Samples
500
1000
5000
10000
D: # Response Levels
2
3
4
E: Distribution
Normal
Gamma
Uniform
Mixed


Correlation Plot For QC Assay vs PAC Data Rank: Ordinal Regression vs GA
QC Assay Data Rank
PAC Data Rank
0
40
80
120
120
111
98
80
40
0
Source
BIT
BO
CGO
FO
HFO
KER
NAPTHA
OGO
P.LAT
RAE
SRGO
TDAE
UATO
UDAE
VHGO
WAX