Determining the Right Sample Size for an MSA Study

Laura Lancaster and Chris Gotwalt

JMP Research & Development
SAS Institute

Discovery Summit Europe 2017

Outline

1. Measurement Systems Analysis
2. Previous Study
3. Current Study Design
4. Results
   - Two Factors Crossed
   - Two Factors Nested
5. Conclusions

Measurement Systems Analysis (MSA)

An MSA study is a designed experiment that helps determine how much measurement variation is contributing to overall process variation.

Measurement Systems Analysis (MSA)

An MSA study is a designed experiment that helps determine how much measurement variation is contributing to overall process variation.

- These studies use random effects models to estimate variance components that assess the sources of variation in the measurement process.

Measurement Systems Analysis (MSA)

An MSA study is a designed experiment that helps determine how much measurement variation is contributing to overall process variation.

- These studies use random effects models to estimate variance components that assess the sources of variation in the measurement process.
- The variance components are typically estimated using one of three methods:
  - Average and Range Method
  - Expected Means Squares (EMS)
  - Restricted Maximum Likelihood (REML)

Measurement Systems Analysis (MSA)

An MSA study is a designed experiment that helps determine how much measurement variation is contributing to overall process variation.

- These studies use random effects models to estimate variance components that assess the sources of variation in the measurement process.
- The variance components are typically estimated using one of three methods:
  - Average and Range Method
  - Expected Means Squares (EMS)
  - Restricted Maximum Likelihood (REML)
- **Problem:** These methods can produce negative variance component estimates that do not make sense in MSA studies.

Measurement Systems Analysis (MSA)

An MSA study is a designed experiment that helps determine how much measurement variation is contributing to overall process variation.

- These studies use random effects models to estimate variance components that assess the sources of variation in the measurement process.
- The variance components are typically estimated using one of three methods:
  - Average and Range Method
  - Expected Means Squares (EMS)
  - Restricted Maximum Likelihood (REML)
- **Problem:** These methods can produce negative variance component estimates that do not make sense in MSA studies.
- **Typical Solution:** Set negative variance components to zero.

Measurement Systems Analysis (MSA)

An MSA study is a designed experiment that helps determine how much measurement variation is contributing to overall process variation.

- These studies use random effects models to estimate variance components that assess the sources of variation in the measurement process.
- The variance components are typically estimated using one of three methods:
  - Average and Range Method
  - Expected Means Squares (EMS)
  - Restricted Maximum Likelihood (REML)
- **Problem:** These methods can produce negative variance component estimates that do not make sense in MSA studies.
- **Typical Solution:** Set negative variance components to zero.
- **Some practitioners were not happy with zeroed variance components either!!**

MSA - Bayesian Estimate Method

New Solution: We found a Bayesian estimation method that produces strictly positive variance components using a non-informative prior.

MSA - Bayesian Estimate Method

New Solution: We found a Bayesian estimation method that produces strictly positive variance components using a non-informative prior.

-
- We generalized Portnoy and Sahai's modified Jeffrey's Prior and implemented it in JMP's Variability platform.

MSA - Bayesian Estimate Method

New Solution: We found a Bayesian estimation method that produces strictly positive variance components using a non-informative prior.

-
- We generalized Portnoy and Sahai's modified Jeffrey's Prior and implemented it in JMP's Variability platform.
- JMP's default behavior is to use REML estimates if no variance components have been zeroed and use the Bayesian estimates otherwise. We will refer to this as a **Hybrid method**.

Previous Study

1. Compared the bias and variability of the Bayesian and Hybrid estimates to the REML estimates.

Previous Study

1. Compared the bias and variability of the Bayesian and Hybrid estimates to the REML estimates.
2. Compared each estimation method's ability to correctly classify measurement systems.

Previous Study

1. Compared the bias and variability of the Bayesian and Hybrid estimates to the REML estimates.
2. Compared each estimation method's ability to correctly classify measurement systems.
   - Don Wheeler's *Evaluating the Measurement Process* (EMP) method

Previous Study

1. Compared the bias and variability of the Bayesian and Hybrid estimates to the REML estimates.
2. Compared each estimation method's ability to correctly classify measurement systems.
   - Don Wheeler's *Evaluating the Measurement Process* (EMP) method
   - Automotive Industry Action Group's (AIAG) *Gauge R&R* method

EMP Classification System

- **Intraclass Correlation Coefficient (ICC)** - ρ - ratio of product variance to total variance

$$\rho = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} = \frac{\sigma_p^2}{\sigma_x^2}$$

EMP Classification System

- **Intraclass Correlation Coefficient (ICC)** - ρ - ratio of product variance to total variance

$$\rho = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} = \frac{\sigma_p^2}{\sigma_x^2}$$

- EMP Classifications:

Classification	$\hat{\rho}$	**Probability of Warning***
First Class	0.80 – 1.00	0.99 – 1.00
Second Class	0.50 – 0.80	0.88 – 0.99
Third Class	0.20 – 0.50	0.40 – 0.88
Fourth Class	0.00 – 0.20	0.03 – 0.40

* Probability of a warning for a $3\sigma_p$ shift within 10 subgroups using Test 1.

AIAG's Classification System

- AIAG uses Percent Gauge R&R to classify the health of a measurement system.

$$\%GRR = 100\sqrt{\frac{\hat{\sigma}_e^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2}} = 100\frac{\hat{\sigma}_e}{\hat{\sigma}_x} = 100\sqrt{1 - \hat{\rho}}$$

AIAG's Classification System

- AIAG uses Percent Gauge R&R to classify the health of a measurement system.

$$\%GRR = 100\sqrt{\frac{\hat{\sigma}_e^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2}} = 100\frac{\hat{\sigma}_e}{\hat{\sigma}_x} = 100\sqrt{1 - \hat{\rho}}$$

- AIAG Classifications:

Classification	$\%GRR$	$\hat{\rho}$
Acceptable	$0\% - 10\%$	$0.99 - 1.00$
Marginal	$10\% - 30\%$	$0.91 - 0.99$
Unacceptable	$30\% - 100\%$	$0.00 - 0.91$

Previous Simulation Study

- Three Typical MSA Designs

Previous Simulation Study

- Three Typical MSA Designs
  - Two Factors Crossed (balanced) with 3 Operators, 10 Parts, 3 Replications

Previous Simulation Study

- Three Typical MSA Designs
  - Two Factors Crossed (balanced) with 3 Operators, 10 Parts, 3 Replications
  - Two Factors Nested (balanced) with 3 Operators, 20 Parts, 2 Replications

Previous Simulation Study

- Three Typical MSA Designs
  - Two Factors Crossed (balanced) with 3 Operators, 10 Parts, 3 Replications
  - Two Factors Nested (balanced) with 3 Operators, 20 Parts, 2 Replications
  - Three Factors Staggered Nested Design (highly unbalanced) with 120 measurements - Performance was very bad!

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.
- **Variability** - Bayesian estimates almost always have smaller RMSE and standard deviations than REML and Hybrid estimates.

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.
- **Variability** - Bayesian estimates almost always have smaller RMSE and standard deviations than REML and Hybrid estimates.
- **EMP classifications**:
  - Were generally not too bad for all methods and designs but we thought they could be improved by increasing sample size.

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.
- **Variability** - Bayesian estimates almost always have smaller RMSE and standard deviations than REML and Hybrid estimates.
- **EMP classifications**:
  - Were generally not too bad for all methods and designs but we thought they could be improved by increasing sample size.
  - Worst case was 40% incorrect classification by all methods for a two factors crossed design with a class 3 system.

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.
- **Variability** - Bayesian estimates almost always have smaller RMSE and standard deviations than REML and Hybrid estimates.
- **EMP classifications**:
  - Were generally not too bad for all methods and designs but we thought they could be improved by increasing sample size.
  - Worst case was 40% incorrect classification by all methods for a two factors crossed design with a class 3 system.
- **AIAG classifications**:
  - Were generally pretty good for two factors crossed but very bad for two factors nested designs, especially for acceptable systems.

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.
- **Variability** - Bayesian estimates almost always have smaller RMSE and standard deviations than REML and Hybrid estimates.
- **EMP classifications**:
  - Were generally not too bad for all methods and designs but we thought they could be improved by increasing sample size.
  - Worst case was 40% incorrect classification by all methods for a two factors crossed design with a class 3 system.
- **AIAG classifications**:
  - Were generally pretty good for two factors crossed but very bad for two factors nested designs, especially for acceptable systems.
  - We hoped that increasing sample size would help with these classifications.

Previous Study Results

- **Bias** - REML and Hybrid estimates are generally less biased than Bayesian estimates.
- **Variability** - Bayesian estimates almost always have smaller RMSE and standard deviations than REML and Hybrid estimates.
- **EMP classifications**:
  - Were generally not too bad for all methods and designs but we thought they could be improved by increasing sample size.
  - Worst case was 40% incorrect classification by all methods for a two factors crossed design with a class 3 system.
- **AIAG classifications**:
  - Were generally pretty good for two factors crossed but very bad for two factors nested designs, especially for acceptable systems.
  - We hoped that increasing sample size would help with these classifications.
  - Worst case was 100% incorrect classification by the Bayesian method for the two factors nested design with an acceptable system.

Current Research - Sample Size

How does sample size affect our ability to estimate the variance components and classify systems with the EMP and AIAG methods?

Simulation Study Design

- Studied 2 Factors Crossed and 2 Factors Nested designs.

Simulation Study Design

- Studied 2 Factors Crossed and 2 Factors Nested designs.
- Range of bad to good measurement systems (using ICC as the metric)

Simulation Study Design

- Studied 2 Factors Crossed and 2 Factors Nested designs.
- Range of bad to good measurement systems (using ICC as the metric)
  - ICC values in middle of EMP classifications: 0.1, 0.35, 0.65, 0.9

Simulation Study Design

- Studied 2 Factors Crossed and 2 Factors Nested designs.
- Range of bad to good measurement systems (using ICC as the metric)
  - ICC values in middle of EMP classifications:
    0.1, 0.35, 0.65, 0.9
  - ICC values in middle of AIAG's top 2 classifications:
    0.96 and 0.9975

Simulation Study Design

- Studied 2 Factors Crossed and 2 Factors Nested designs.
- Range of bad to good measurement systems (using ICC as the metric)
  - ICC values in middle of EMP classifications: 0.1, 0.35, 0.65, 0.9
  - ICC values in middle of AIAG's top 2 classifications: 0.96 and 0.9975
- Part variance value: 5

Simulation Study Design

- Studied 2 Factors Crossed and 2 Factors Nested designs.
- Range of bad to good measurement systems (using ICC as the metric)
  - ICC values in middle of EMP classifications: 0.1, 0.35, 0.65, 0.9
  - ICC values in middle of AIAG's top 2 classifications: 0.96 and 0.9975
- Part variance value: 5
- 250 Simulations

Simulation Study Design

- We used JSL in JMP Pro 13 to run the simulations.

Simulation Study Design

- We used JSL in JMP Pro 13 to run the simulations.
- Used the new JMP Pro 13 Simulate function that makes simulating statistics in a JMP report very easy.

Simulation Study Design

- We used JSL in JMP Pro 13 to run the simulations.
- Used the new JMP Pro 13 Simulate function that makes simulating statistics in a JMP report very easy.
- Called the following estimation methods in the Variability platform:

Simulation Study Design

- We used JSL in JMP Pro 13 to run the simulations.
- Used the new JMP Pro 13 Simulate function that makes simulating statistics in a JMP report very easy.
- Called the following estimation methods in the Variability platform:
  - REML

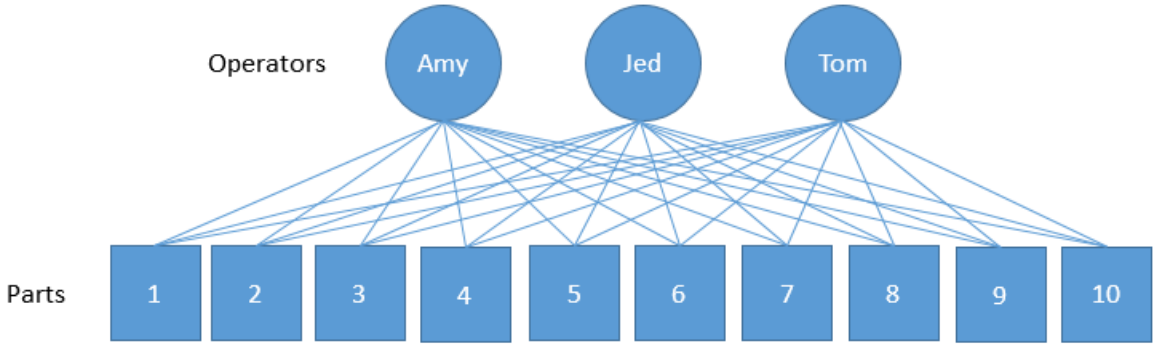
Simulation Study Design

- We used JSL in JMP Pro 13 to run the simulations.
- Used the new JMP Pro 13 Simulate function that makes simulating statistics in a JMP report very easy.
- Called the following estimation methods in the Variability platform:
  - REML
  - Bayesian (Portnoy-Sahai)

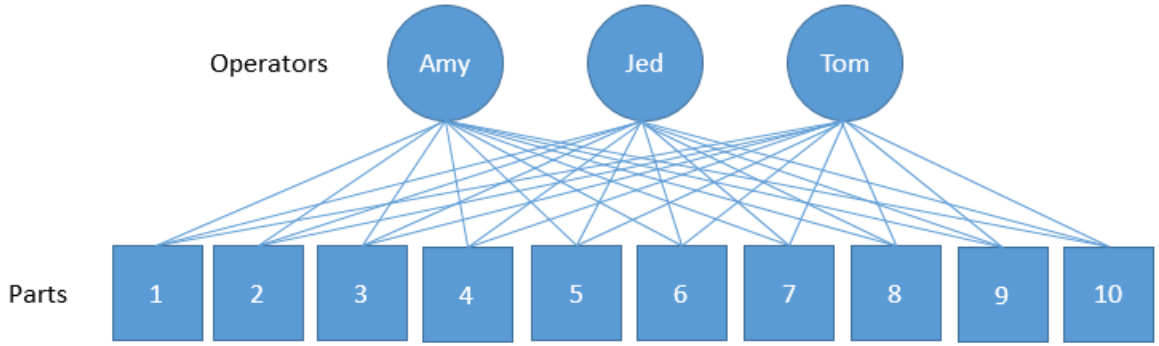
Simulation Study Design

- We used JSL in JMP Pro 13 to run the simulations.
- Used the new JMP Pro 13 Simulate function that makes simulating statistics in a JMP report very easy.
- Called the following estimation methods in the Variability platform:
  - REML
  - Bayesian (Portnoy-Sahai)
  - Hybrid (JMP default setting)
    If zeroed variance components $\Rightarrow$ *Bayesian estimates.*
    Otherwise $\Rightarrow$ *REML estimates.*

Two Factors Crossed Design

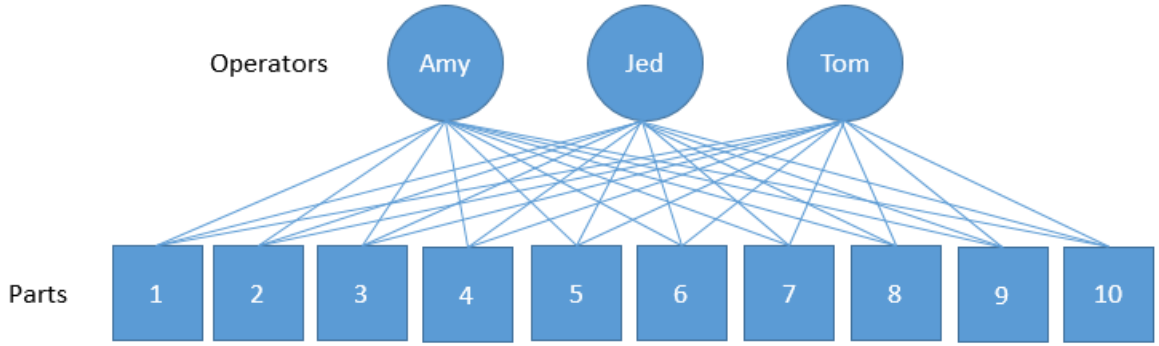


Two Factors Crossed Design



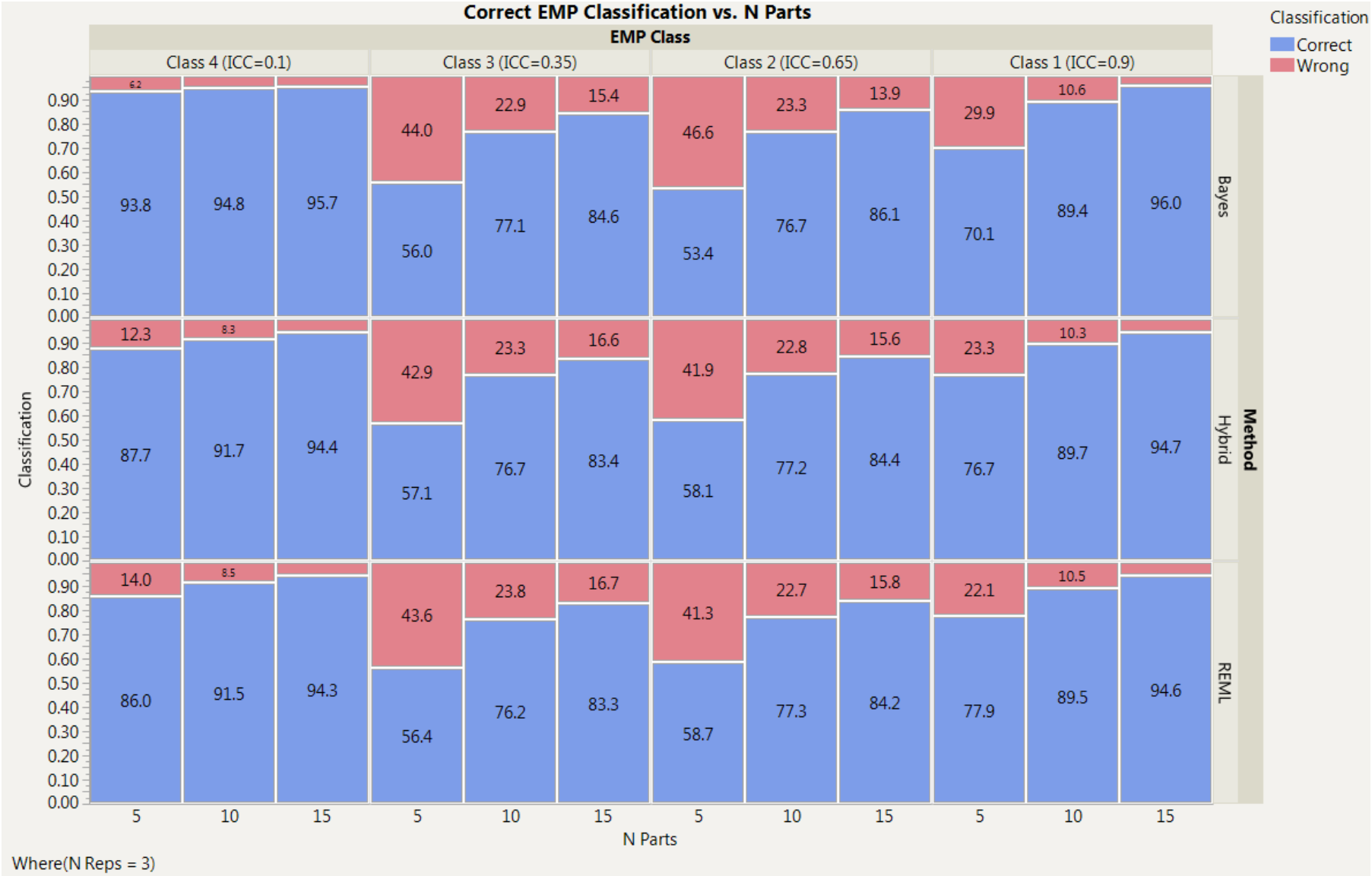
- Balanced design:
  - Number of Operators: 3, 6, 9, 12
  - Number of Parts: 5, 10, 15
  - Number of Replications: 2, 3

Two Factors Crossed Design

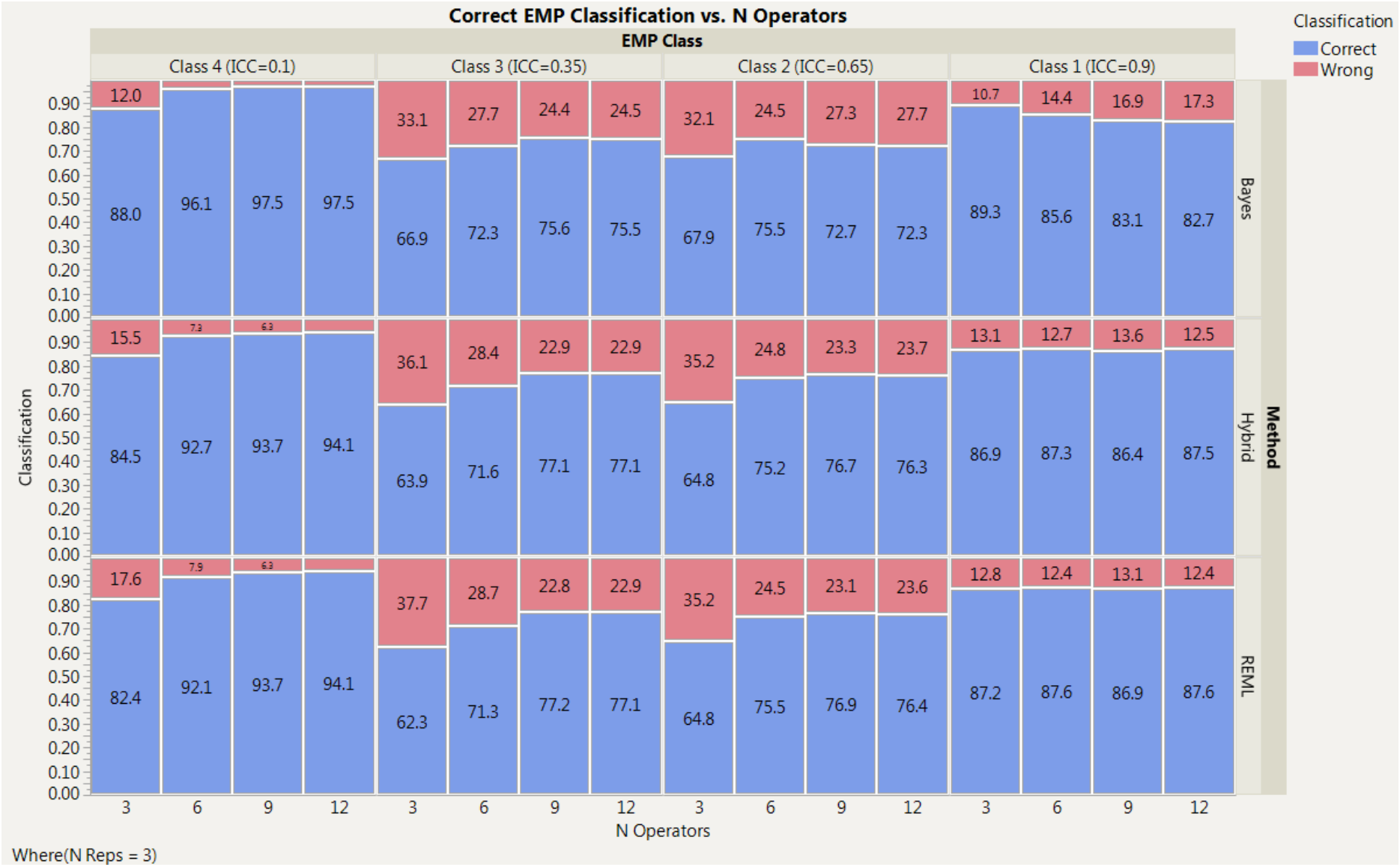


- Balanced design:
  - Number of Operators: 3, 6, 9, 12
  - Number of Parts: 5, 10, 15
  - Number of Replications: 2, 3
- Error variance breakdown:
  - Operator variance = $0.45^{*}\sigma_e^2$
  - Operator*Part variance = $0.1^{*}\sigma_e^2$
  - Residual variance = $0.45^{*}\sigma_e^2$

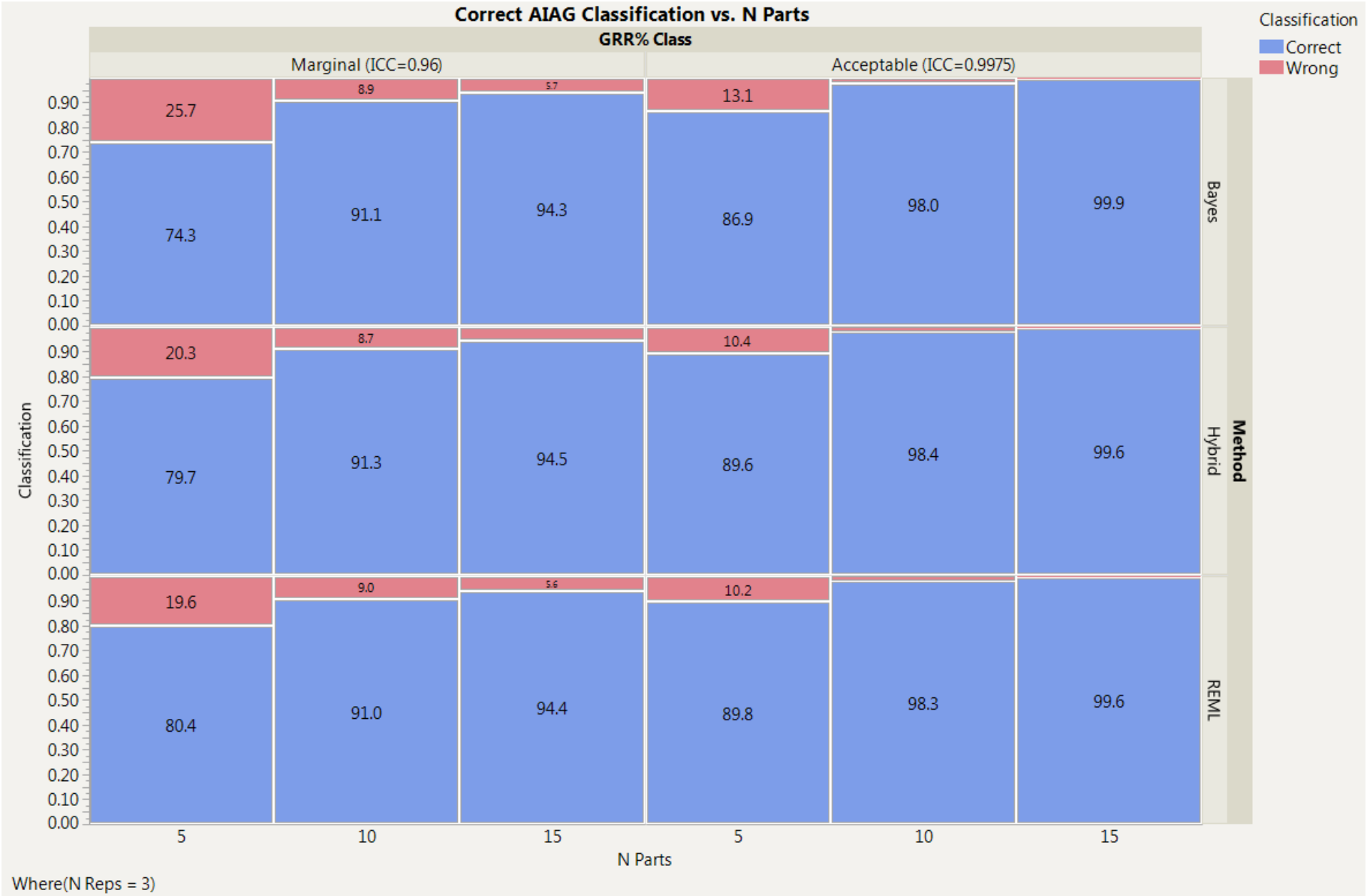
Two Factors Crossed - EMP Classifications



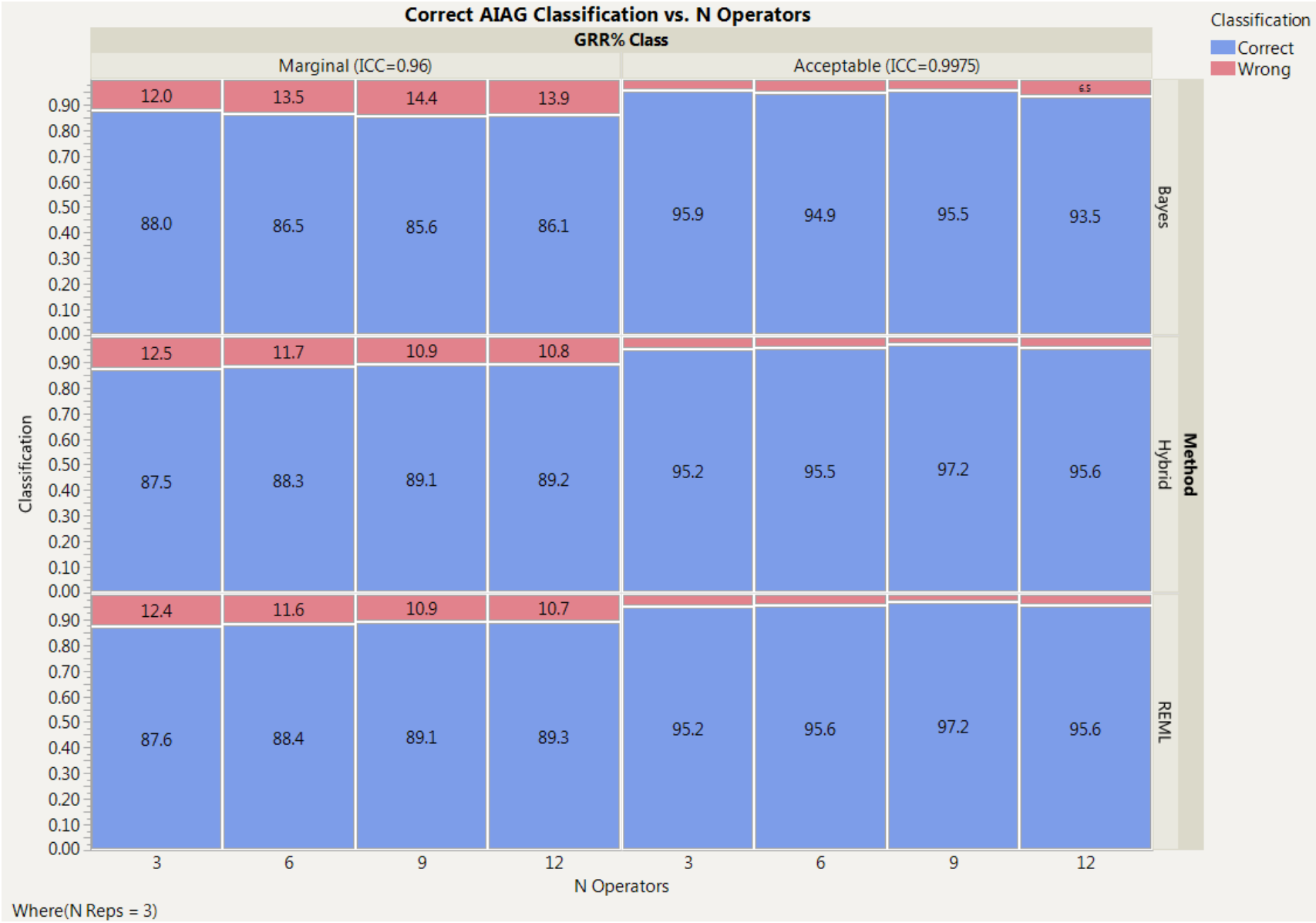
Two Factors Crossed - EMP Classifications



Two Factors Crossed - AIAG Classifications



Two Factors Crossed - AIAG Classifications



Two Factors Crossed - Summary

- EMP Classifications
  - All methods are correct about the same amount.

Two Factors Crossed - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most except for really bad systems (class 4).

Two Factors Crossed - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most except for really bad systems (class 4).
- AIAG Classifications
  - All methods are correct about the same amount.

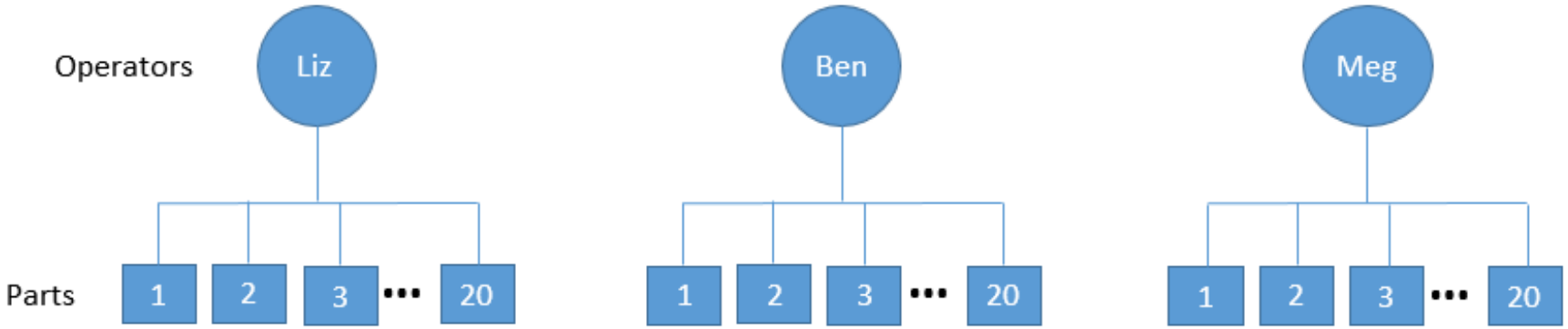
Two Factors Crossed - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most except for really bad systems (class 4).
- AIAG Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most. Increasing operators does not have much impact.

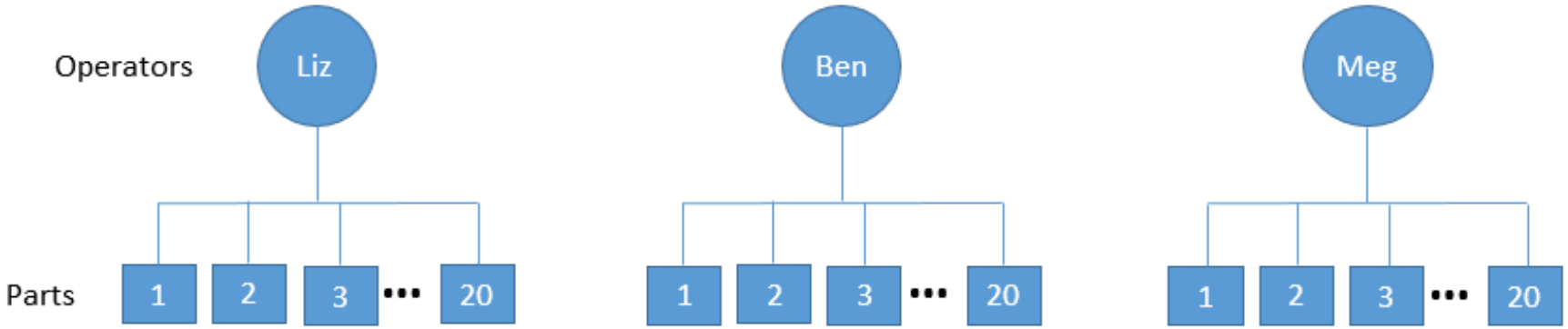
Two Factors Crossed - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most except for really bad systems (class 4).
- AIAG Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most. Increasing operators does not have much impact.
- **Recommendation: Use more than 3 operators (especially for EMP classifications) and at least 10 parts.**

Two Factors Nested Design

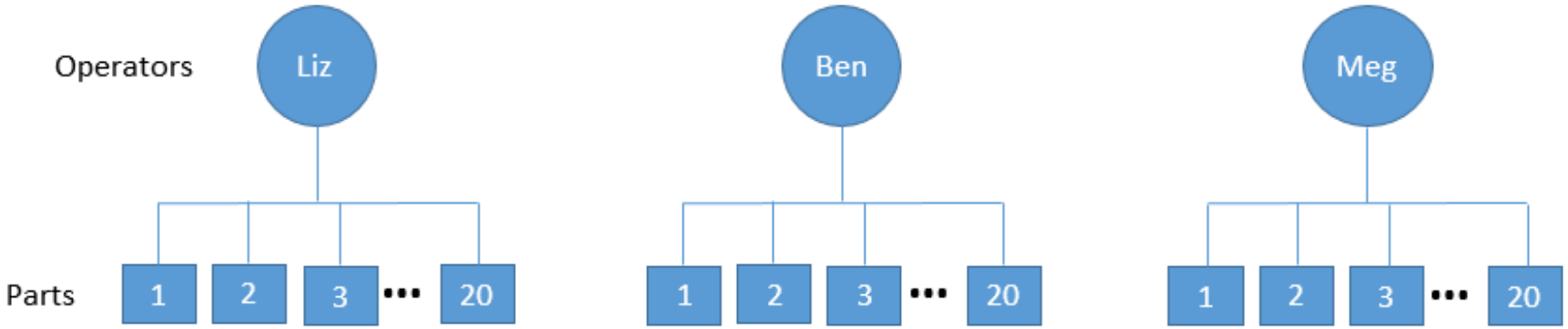


Two Factors Nested Design



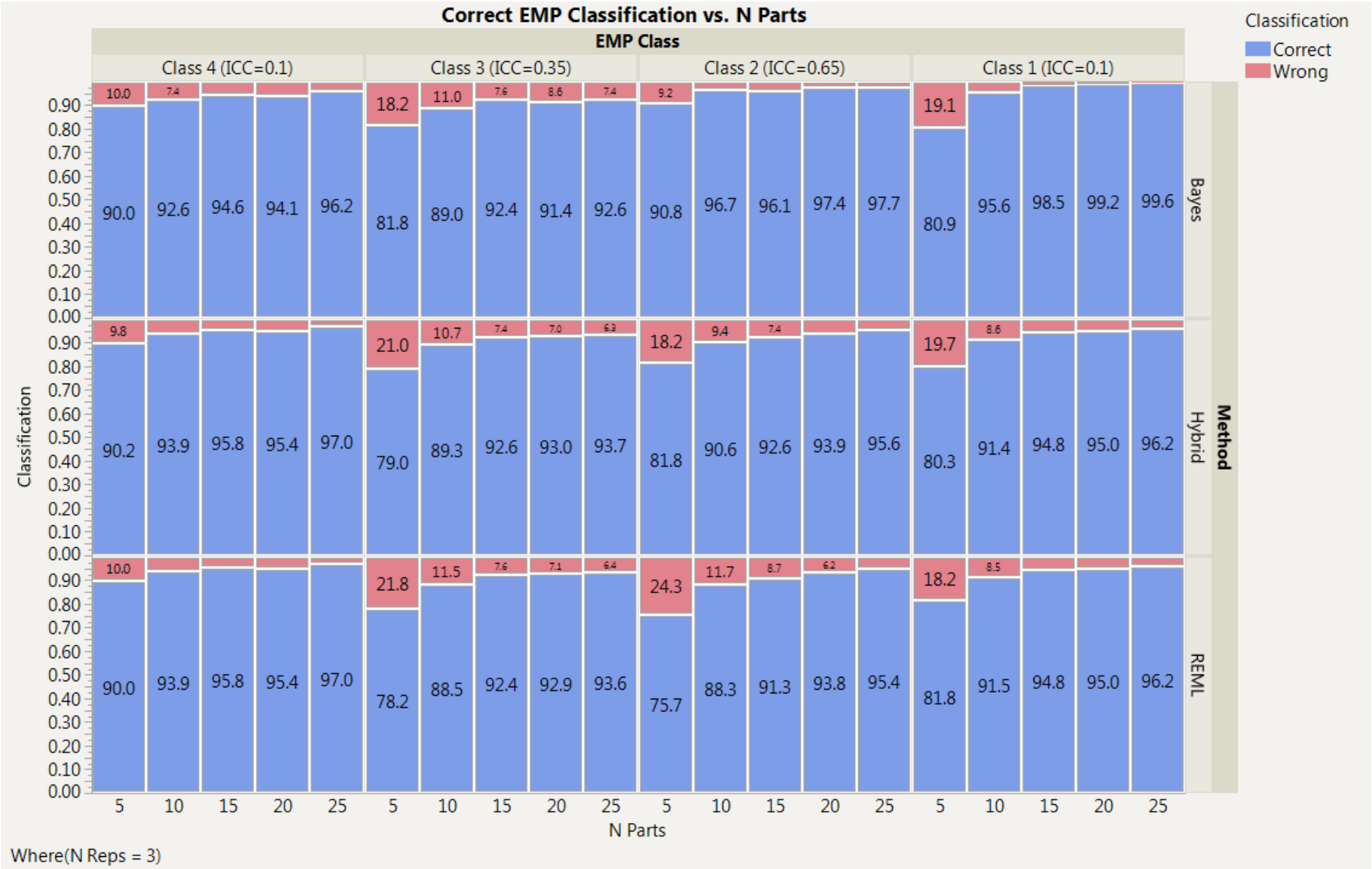
- Balanced Design:
  - Number of Operators: 3, 6, 9, 12, 15
  - Number of Parts: 5, 10, 15, 20, 25
  - Number of Replications: 2, 3

Two Factors Nested Design

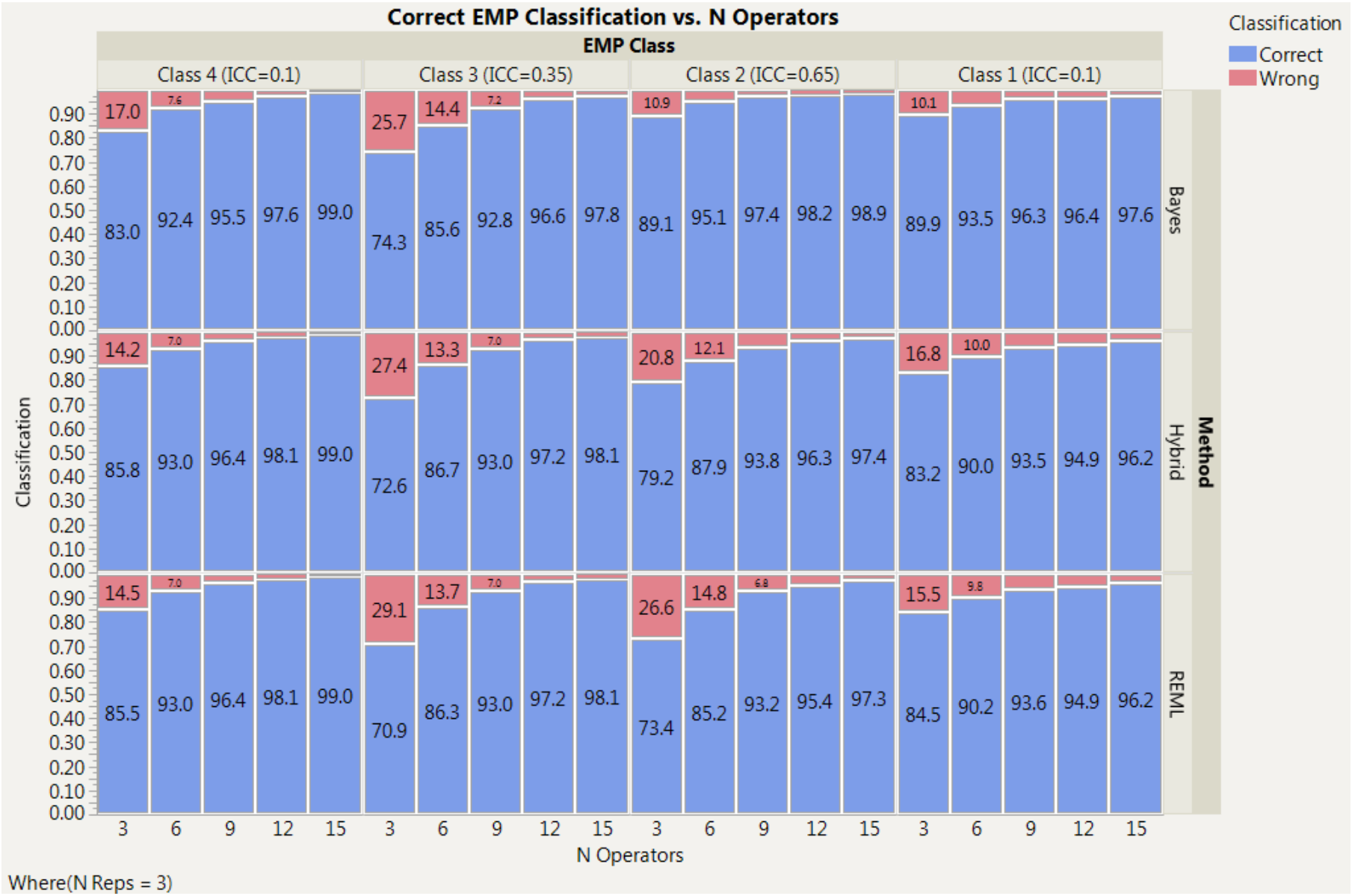


- Balanced Design:
  - Number of Operators: 3, 6, 9, 12, 15
  - Number of Parts: 5, 10, 15, 20, 25
  - Number of Replications: 2, 3
- Error variance breakdown:
  - Operator variance = $0.5^{*}\sigma_e^2$
  - Residual variance = $0.5^{*}\sigma_e^2$

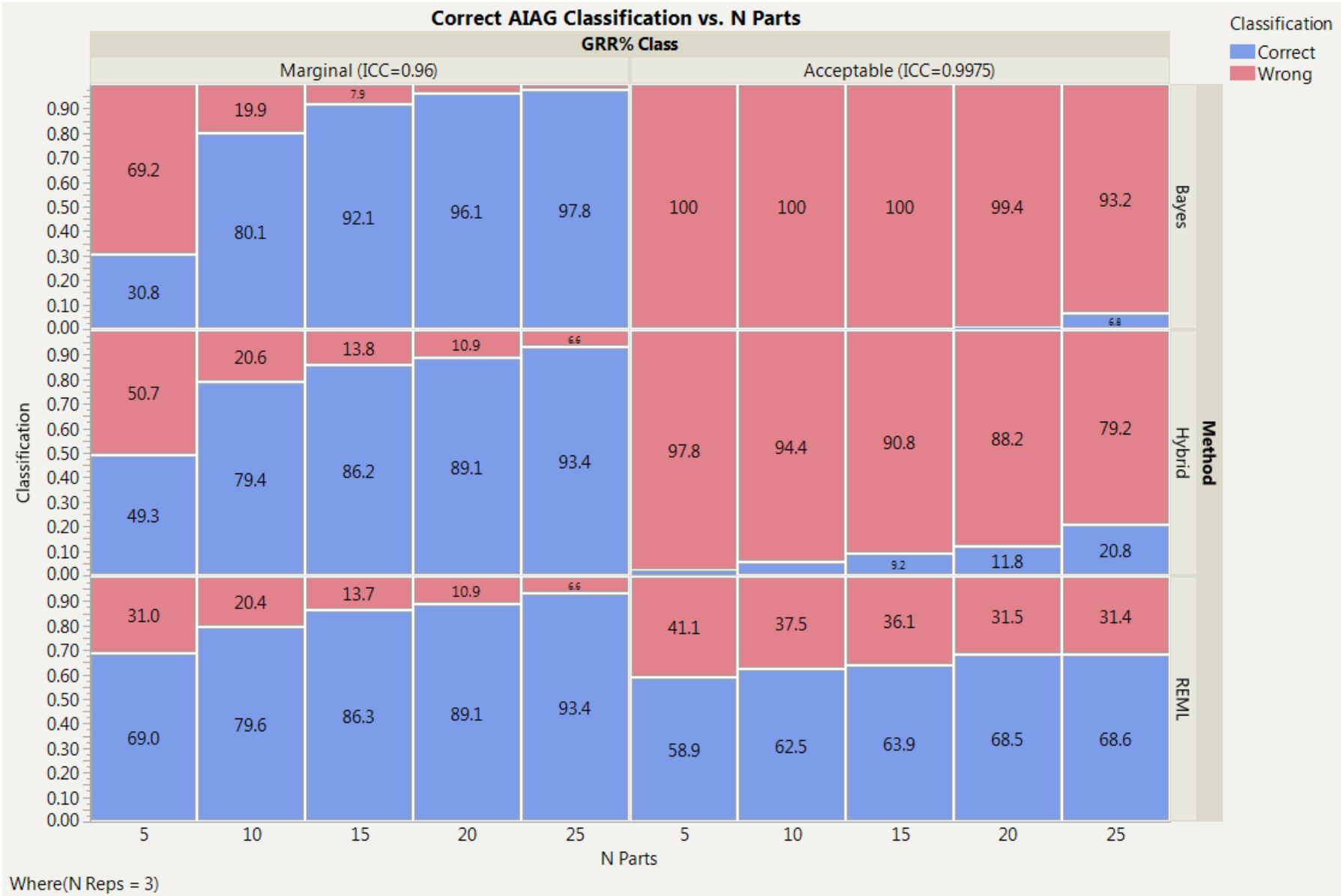
Two Factors Nested - EMP Classifications



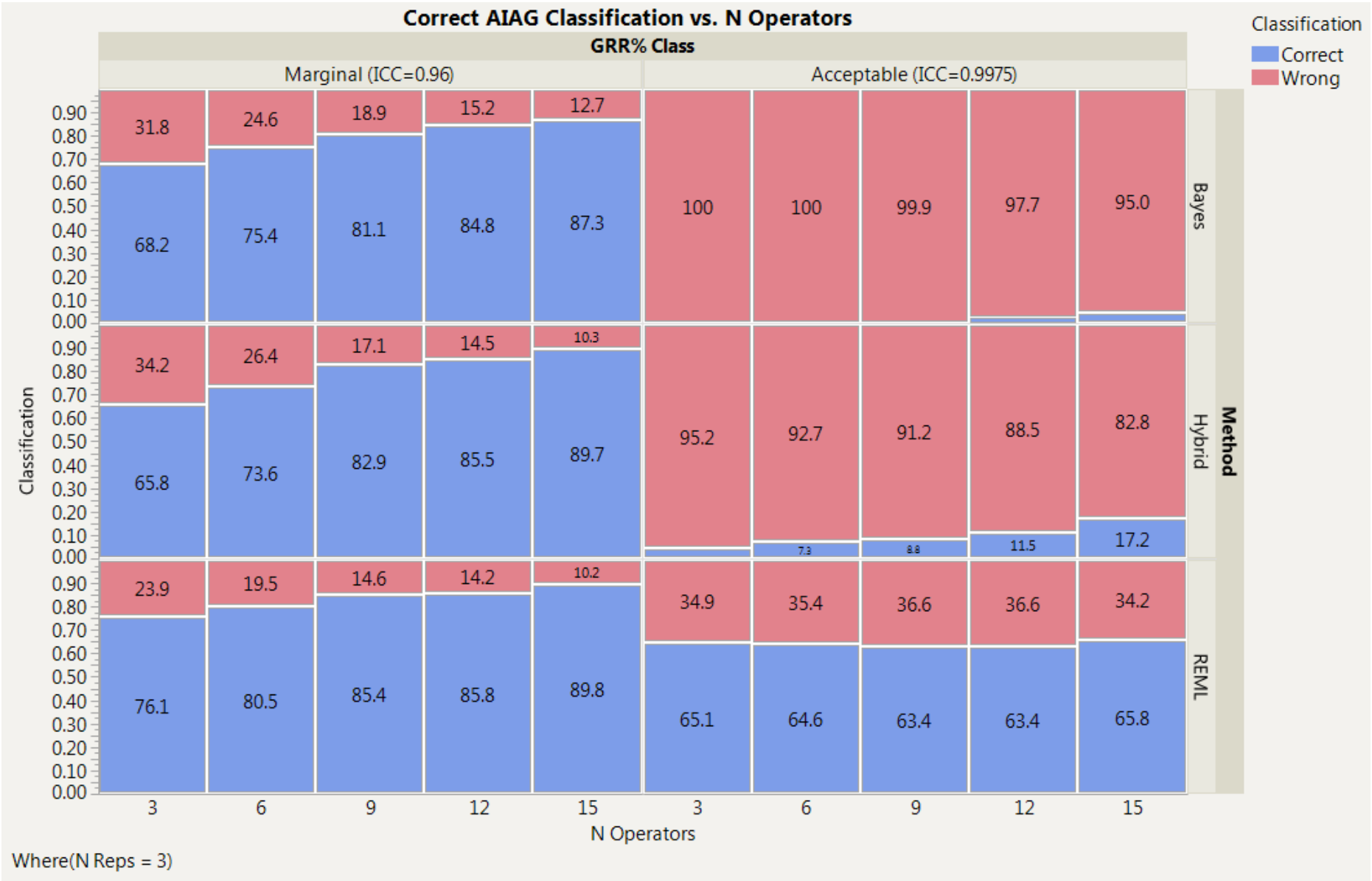
Two Factors Nested - EMP Classifications



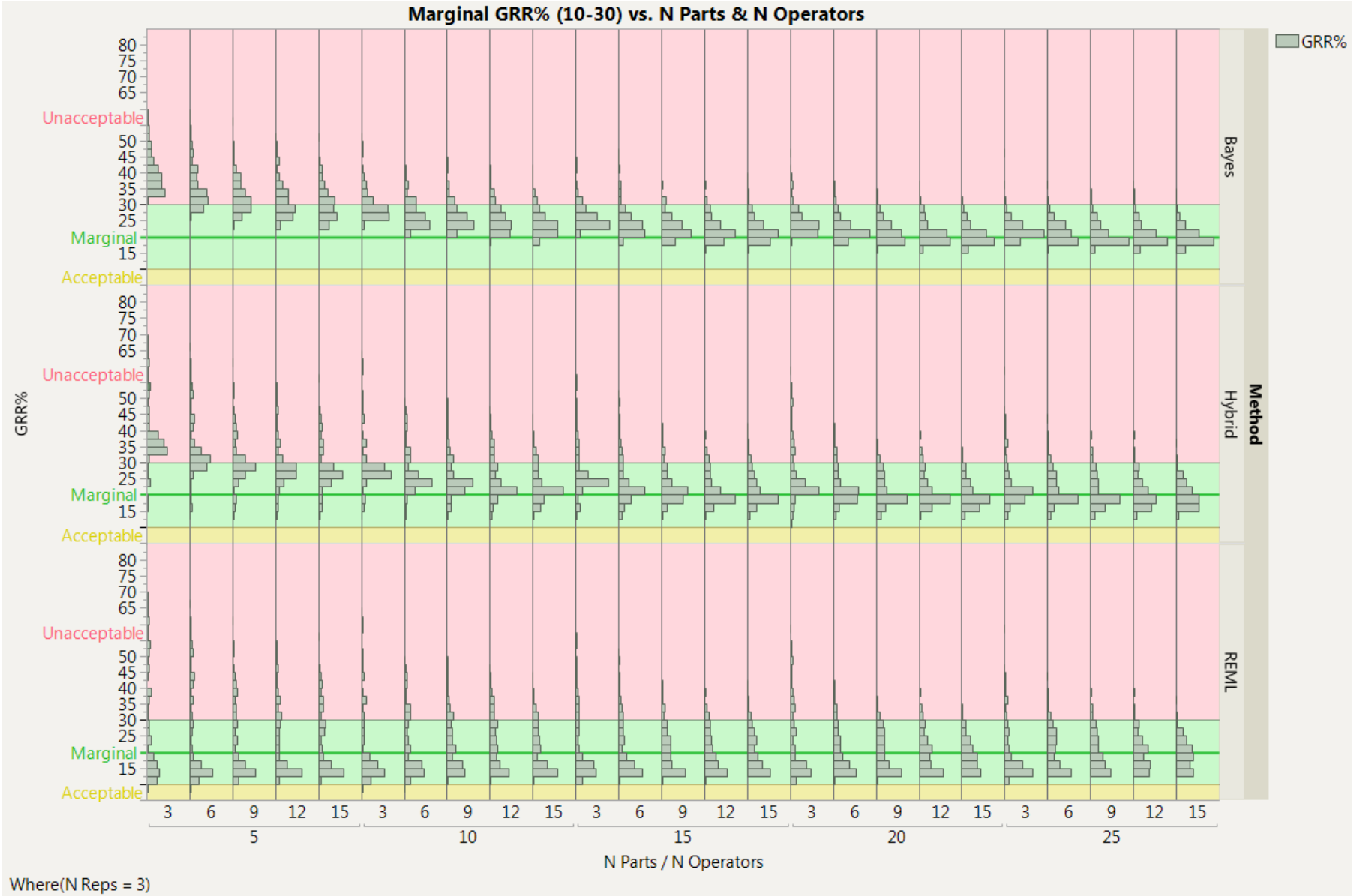
Two Factors Nested - AIAG Classifications



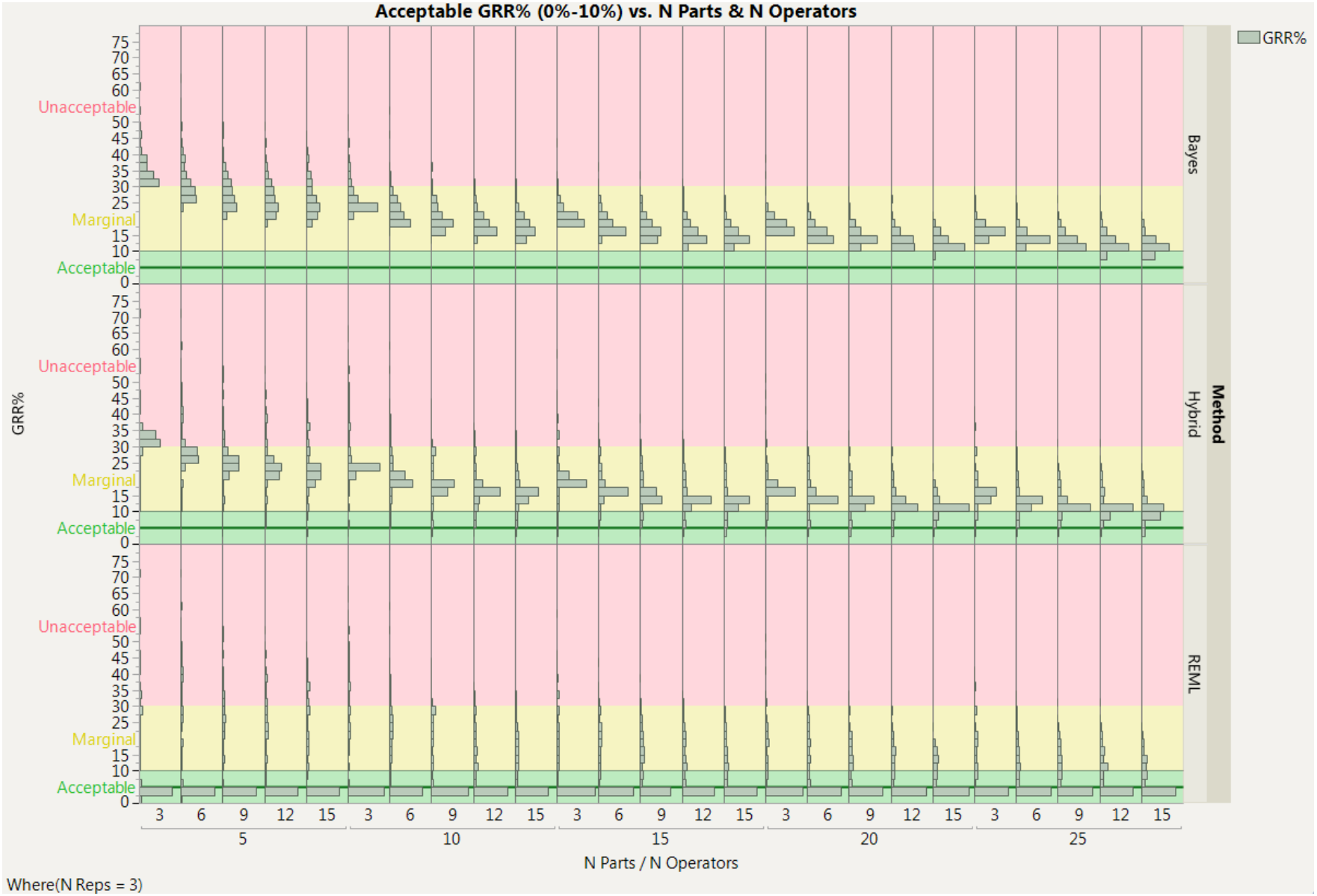
Two Factors Nested - AIAG Classifications



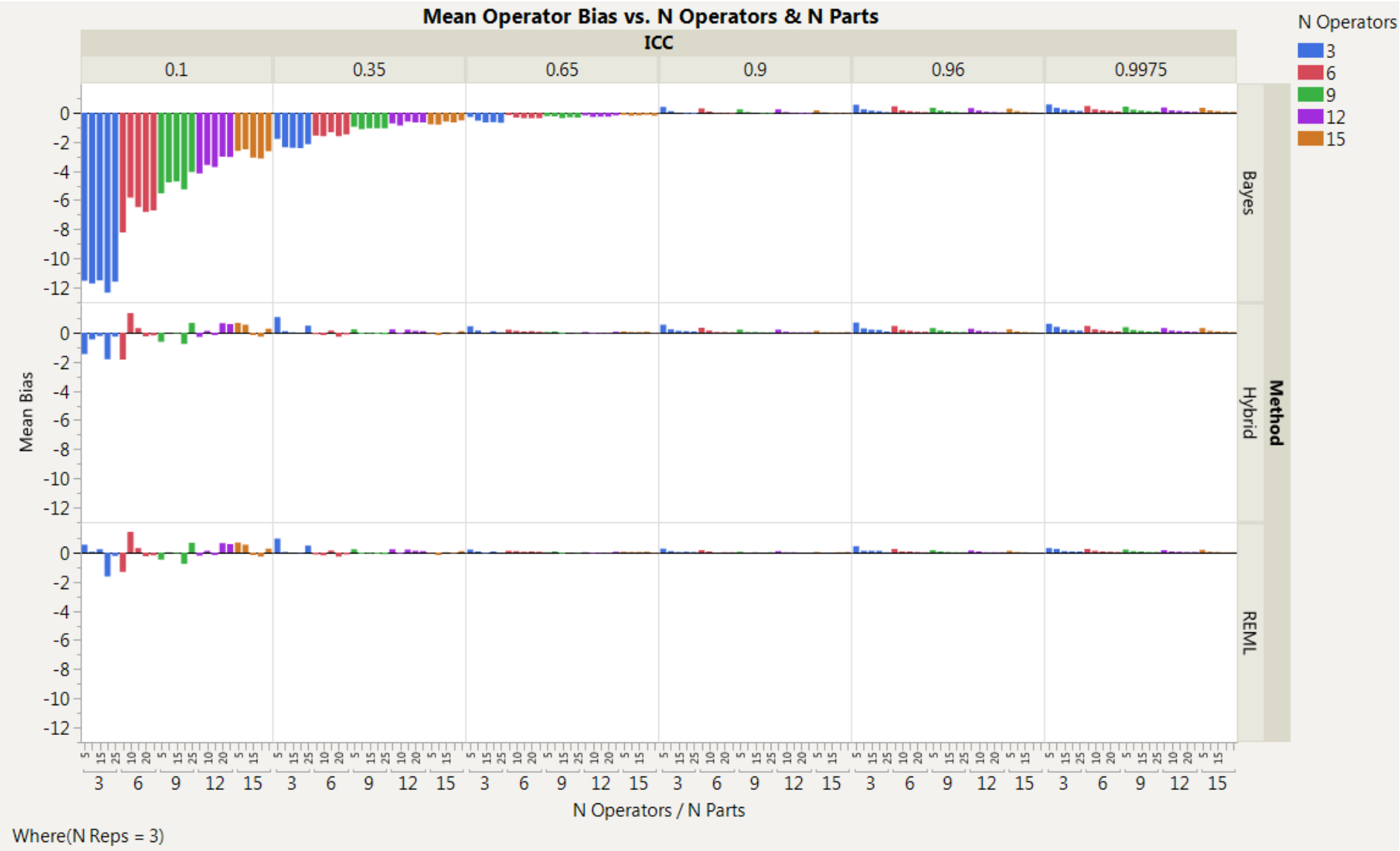
Two Factors Nested - Marginal AIAG Classifications



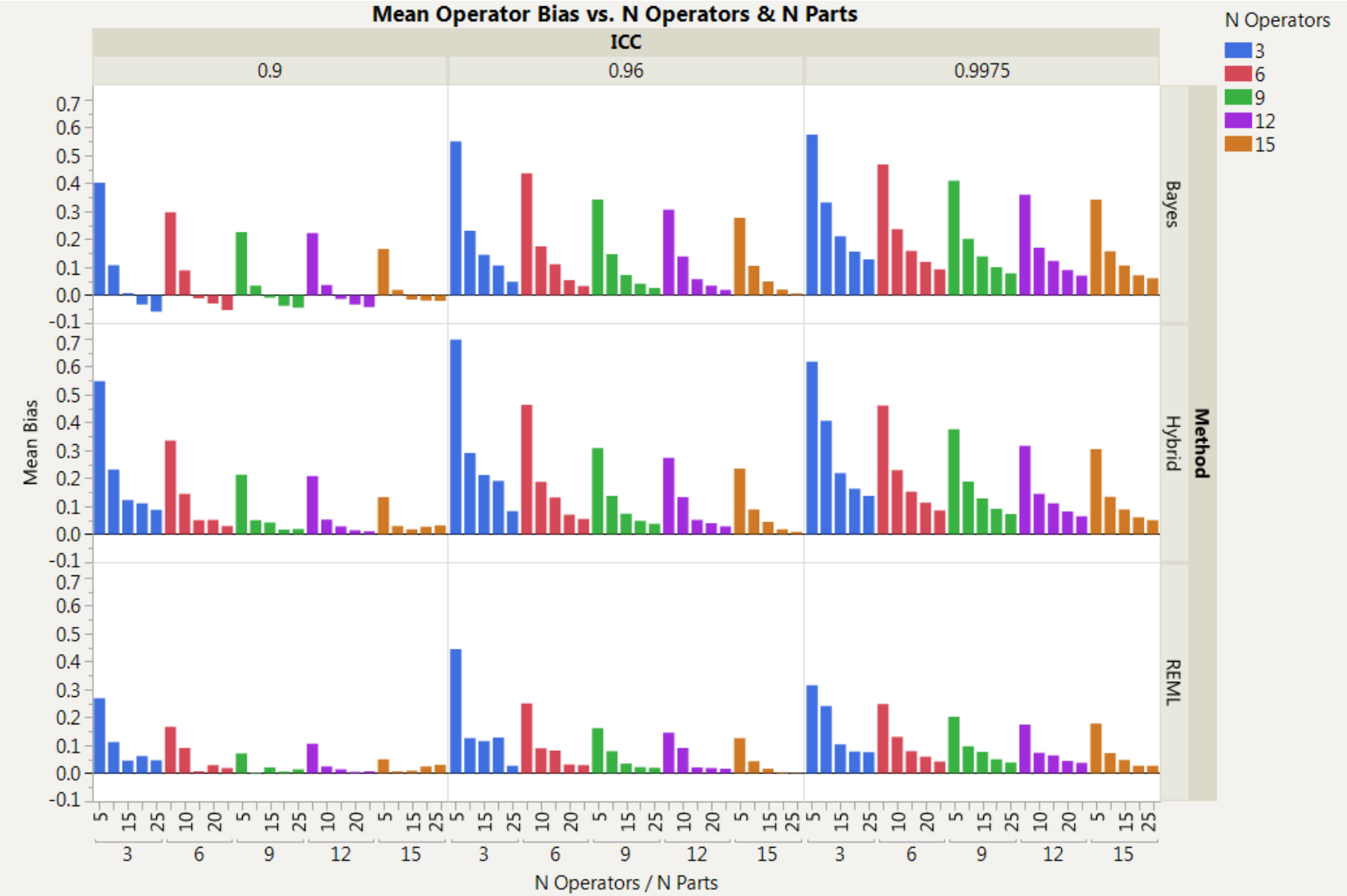
Two Factors Nested - Acceptable AIAG Classifications



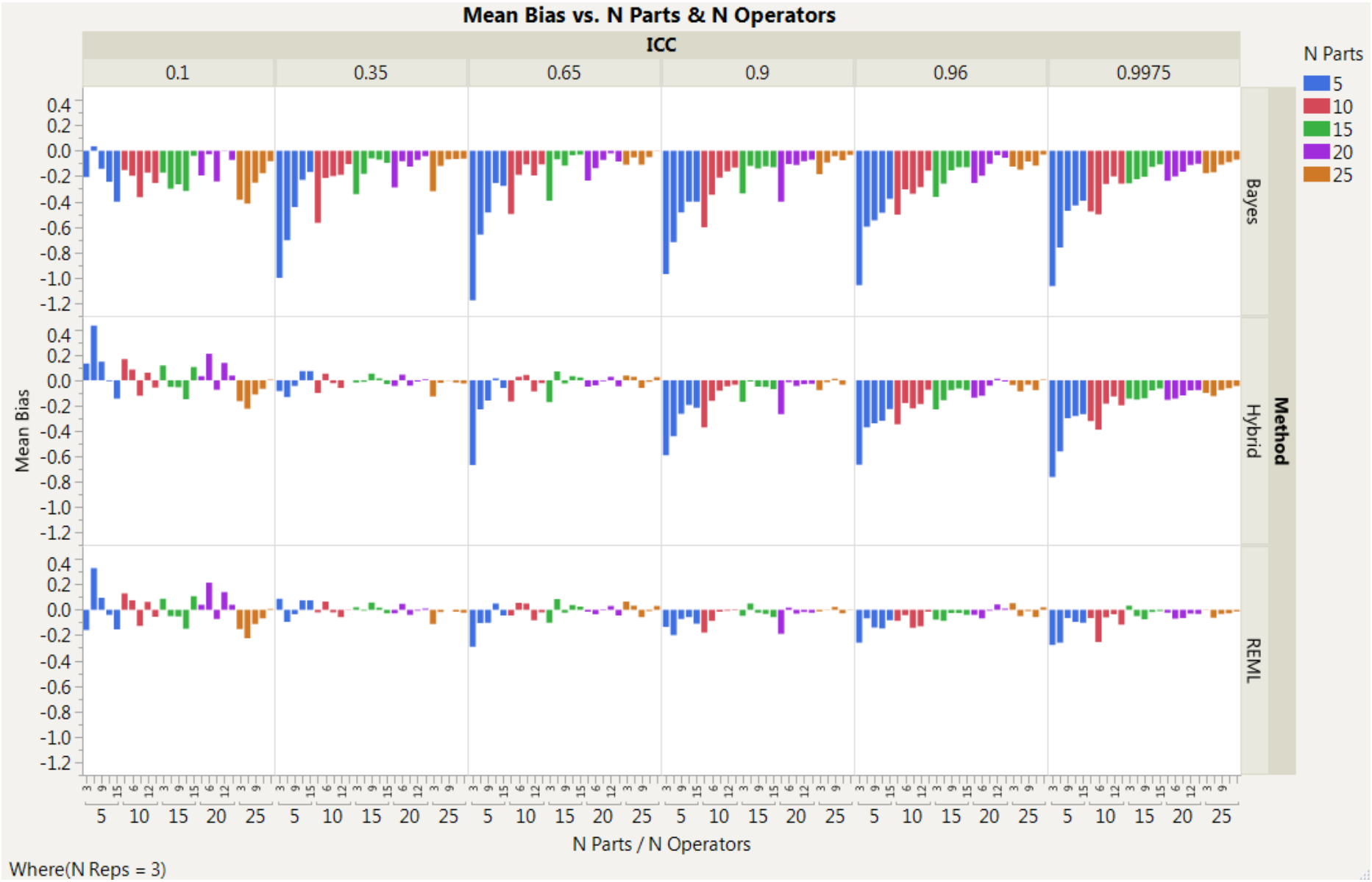
Two Factors Nested - Mean Operator Variance Bias



Two Factors Nested - Mean Operator Variance Bias (Zoom)



Two Factors Nested - Mean Part Variance Bias



Two Factors Nested - Summary

- EMP Classifications
  - All methods are correct about the same amount.

Two Factors Nested - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most for good systems (classes 1 and 2) and increasing operators helps the most with bad systems (classes 3 and 4).

Two Factors Nested - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most for good systems (classes 1 and 2) and increasing operators helps the most with bad systems (classes 3 and 4).
  - **Recommendation: Use more than 3 operators and at least 10 parts.**

Two Factors Nested - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most for good systems (classes 1 and 2) and increasing operators helps the most with bad systems (classes 3 and 4).
  - **Recommendation: Use more than 3 operators and at least 10 parts.**
- AIAG Classifications
  - REML performs the best and is far superior for acceptable systems. (Bayesian and Hybrid do well for marginal systems if you have higher sample sizes but horribly for acceptable systems.)

Two Factors Nested - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most for good systems (classes 1 and 2) and increasing operators helps the most with bad systems (classes 3 and 4).
  - **Recommendation: Use more than 3 operators and at least 10 parts.**
- AIAG Classifications
  - REML performs the best and is far superior for acceptable systems. (Bayesian and Hybrid do well for marginal systems if you have higher sample sizes but horribly for acceptable systems.)
  - Increasing number of parts helps the most. It has more impact for marginal systems.

Two Factors Nested - Summary

- EMP Classifications
  - All methods are correct about the same amount.
  - Increasing number of parts helps the most for good systems (classes 1 and 2) and increasing operators helps the most with bad systems (classes 3 and 4).
  - **Recommendation: Use more than 3 operators and at least 10 parts.**
- AIAG Classifications
  - REML performs the best and is far superior for acceptable systems. (Bayesian and Hybrid do well for marginal systems if you have higher sample sizes but horribly for acceptable systems.)
  - Increasing number of parts helps the most. It has more impact for marginal systems.
  - **Recommendation: Use REML, especially if you think your system is acceptable! Sample sizes with more than 3 operators and at least 10 parts are best. Caution: REML was still only correct 73.2% with 15 operators and 25 parts.**

Conclusions

- **2 Factors Crossed:** Sample sizes of typical 2 factors crossed designs seem to be OK with both classification systems.

Conclusions

- **2 Factors Crossed:** Sample sizes of typical 2 factors crossed designs seem to be OK with both classification systems.
- **2 Factors Nested:** Sample sizes of typical 2 factors nested designs do OK for EMP classifications but not AIAG classifications, especially for acceptable systems.

Future Research

- Fine tune the sample sizes between 3 and 6 operators and 5 and 10 parts.

Future Research

- Fine tune the sample sizes between 3 and 6 operators and 5 and 10 parts.
- Study more types of MSA designs.

Future Research

- Fine tune the sample sizes between 3 and 6 operators and 5 and 10 parts.
- Study more types of MSA designs.
- Try different breakdowns of error variance.

References

- Automotive Industry Action Group (2002), *Measurement Systems Analysis Reference Manual, 3rd Edition*.
- Portnoy (1971), “Formal Bayes Estimation With Application To a Random Effect Model,” *Annals Of Mathematical Statistics*, 42, 1379-1402.
- Sahai (1974), “Some Formal Bayes Estimators of Variance Components in Balanced Three-Stage Nested Random Effects Model,” *Communications in Statistics*, 3, 233-242.

Laura.Lancaster@jmp.com
ChristopherM.Gotwalt@jmp.com