Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

BRAWLER TO CFAM: INCORPORATING STOCHASTIC ENGAGEMENT-LEVEL DATA IN DETERMINISTIC CAMPAIGN MODELS

Benjamin R. Mayo
Todd J. Paciencia
Daniel P. Croghan

Headquarters, Air Force Studies,
Analyses, and Assessments (AF/A9)
Pentagon
1570 Air Force Pentagon
Washington, DC, 20330, USA

ABSTRACT

Headquarters Air Force Studies, Analyses, and Assessments (AF/A9) supports Force Structure decisions by integrating analysis at various levels of resolution. The Combat Forces Assessment Model (CFAM), is a mixed integer program incorporating results from higher-resolution models to identify an optimal force mix within Air Force resources. CFAM is a deterministic model, but some input models are stochastic, such as the tactical air combat simulation BRAWLER. Distributional information is lost when transferring output from BRAWLER as input for CFAM as point estimates. These problems cannot be solved by standard comparison techniques (e.g. Tukey's, Fisher's) because they assume normally distributed data (which BRAWLER data does not satisfy) and both are "overwhelmed" by large number of comparisons. Combining bootstrapping techniques with clustering methods, AF/A9 has created point estimates for CFAM input data, maintaining data integrity. This presentation describes the initial analysis and techniques for using this process in other stochastic-to-deterministic model integrations.

1 INTRODUCTION

Simulating large systems and projects at high resolution poses a number of problems. Increasing model complexity in even the smallest models can have significant impacts on run time and data requirements. To overcome this, smaller, higher resolution models are often used to provide data to simpler models that have a larger scope. This process is used at Headquarters Air Force Studies, Analyses, and Assessments (AF/A9) to supports Force Structure decisions through the integration of engagement, mission, and campaign-level analysis. Engagement-level models have a small scope and very high resolution. One example is BRAWLER, a tactical combat air simulation, uses complex physics modeling in one-on-one or many-versus-many air combat simulations. Analysis from models like BRAWLER is the incorporated in larger models that contain multiple engagements across an entire campaign. One such model is the Combat Forces Assessment Model (CFAM) – a mixed-integer program that is used to identify an optimal mix of assets within projected or unconstrained Air Force Resources. At each stage of analysis transfer – engagement to mission, mission to campaign, etc. – it is accepted that some information is lost; after all, "All models are wrong, some are useful." However, one area for improvement is the incorporation of distributional data in deterministic models. BRAWLER is a stochastic model – force mix options (cases) are given as inputs, with multiple responses used to measure their effectiveness. When this data is input into CFAM, however, only point estimates are used in the mixed-integer program. This means that when

978-1-5090-4486-3/16/$31.00 ©2016 IEEE

two cases have similar distributions but computationally different means, they are considered different under CFAM's parameters. Conversely, if two cases have the same calculated mean, but statistically different distributions, CFAM will treat them as having the same point estimates. See Figures 1 and 2 for comparison.

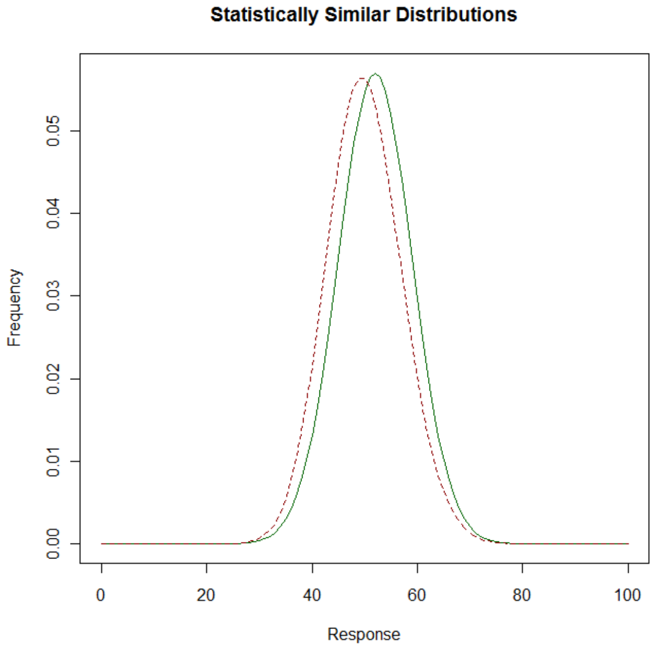

Figure 1: These cases would be treated as different in CFAM.

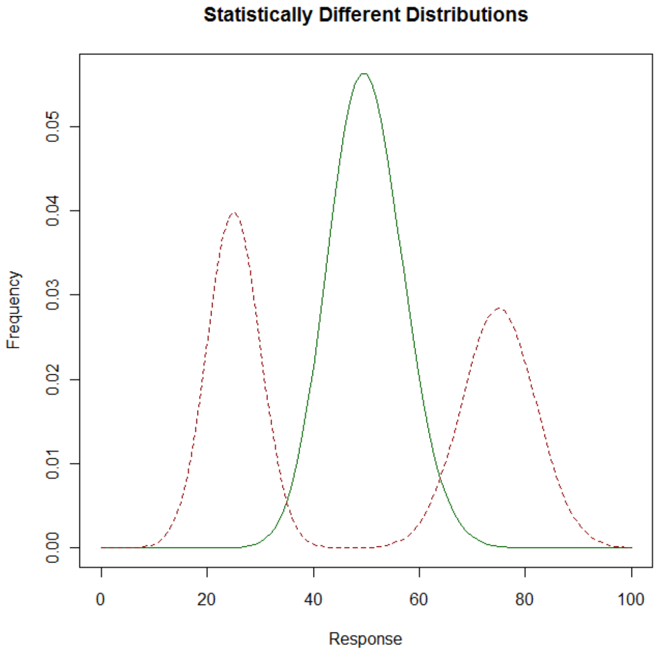

Figure 2: These cases would be treated as the same in CFAM.

2 COMPARING DISTRIBUTIONS

2.1 Bonferroni, Tukey, Fisher Comparisons

The first step in mitigating these problems was to determine a methodology for comparing each case against the others. This allows statistically similar cases to be grouped together so that their point estimates will be the same when input into CFAM. One significant problem with the data is that for some studies the number of cases can be as high as 216. This means that there would be 23,220 pairwise comparisons necessary. One method for conducting multiple comparisons is the Bonferroni correction for confidence intervals, which accounts for the uncertainty across all comparisons (Montgomery, Peck, Vining, 2015):

$$\bar{x} \pm t_{\left(\alpha/_{2p}, df\right)} se(\bar{x})$$

In this formula for the 100(1-α)% confidence interval about the mean, $\bar{x}$, the usual α-value is divided by twice the number of comparisons *(p)*. With 216 individual cases, this would mean that the α-value for an individual confidence interval would be $2.15 * 10^{-6}$, resulting in inordinately wide confidence intervals. Two other often-used comparison methods, Tukey's Honestly Significant Difference (HSD) and Fisher's Least Significant Difference (LSD) (Montgomery, 2008) also fail to differentiate between cases because they assume normally-distributed data. This means that Tukey's and Fisher's will be useful for cases like those in Figure 1, but will fail to differentiate the cases in Figure 2. Additionally, if a case has sufficiently high variance, even if the data is normally-distributed (or at least normal-esque), neither test can differentiate that case from others, as seen in Figure 3.

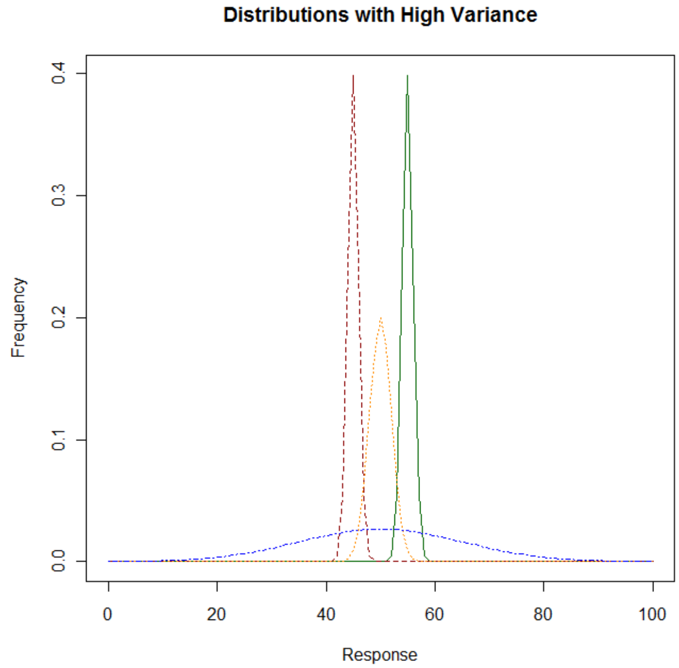

Figure 3: All four distributions are considered similar by Tukey's and Fisher's.

2.2 Bootstrapping

The next method considered was bootstrapping. Bootstrapping is a method of resampling the data (with replacement) and is useful for providing comparisons for when it cannot be assumed that the data is normally-distributed (Mooney, Duval, 1993). Repeated "microsamples" of the original data are taken, the means of which are used to create a confidence interval (Montgomery, Peck, Vining, 2015). Typical runs from BRAWLER have 100 replications for each case. These 100-run samples were used for bootstrapping code implemented in R. Each case was resampled 10,000 times, creating "microsamples" of 100. When the ith sample is taken, the mean, $\hat{x}_i$, for that sample is calculated. Once all 10,000 microsamples are calculated, the simplest method for developing a confidence interval is taking the central $(1-\alpha) * 100\%$ of the microsamples (Montgomery, Peck, Vining, 2015). These bootstrap intervals represent the uncertainty about the mean for each case. Because of this, if there is any overlap between the bootstrap intervals, we cannot assume that the means are different. Bootstrapping proved to be effective in estimating the case means, as well as providing a method for comparing cases against each other.

3 CASE GROUPINGS

3.1 Variance – Based Methods

Simply determining those cases that have pairwise statistical similarities is insufficient for accommodating distributional data into the mixed-integer program. Consider the case in Figure 4 below:

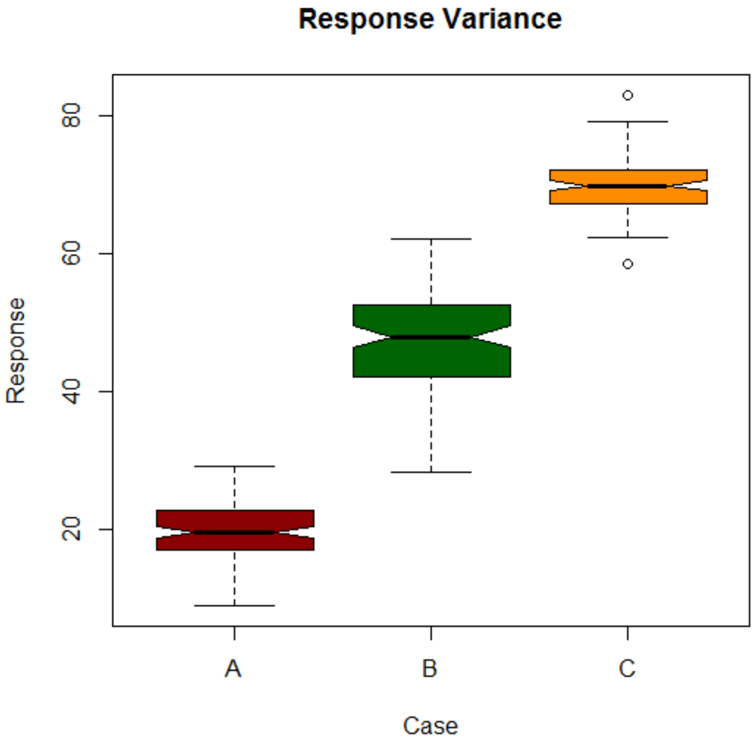

Figure 4: Case A ≈ Case B, Case B ≈ Case C, Case A ≉ Case C.

Case A and Case B have overlapping bootstrap intervals; Case B and Case C have overlapping bootstrap intervals, but there is no overlap between Case A and Case C. The question now becomes how should these cases be grouped when transferring data to CFAM? The initial attempt was simple: define cluster i

as all other cases that are statistically similar to case i (up to case 216 in our example). Calculate the variance of the original means for each case in cluster *i.* Determine the cluster with the lowest variance, which will be considered a group of statistically similar cases. Remove all cases in cluster i from all other clusters, which ensures all groups will be mutually exclusive. The mean for group i is the average of the means for all cases in group i. Other options were considered, such taking the centroid of the confidence intervals in group *i,* but averaging the means was the simplest. This action was repeated as long as clusters with more than 2 cases exist.

.

3.2 Clustering Methods

In addition to this simple variance-based technique, two clustering methods were considered. The first was Affinity Propagation Clustering (AP-clustering). AP-clustering iteratively focuses on determining "exemplars" in clusters of data and determining the clusters that best fit with potential exemplars (Frey, Dueck, 2007). AP-clustering uses is based off of a similarity matrix, fortunately, the bootstrap intervals provided an excellent metric for similarity: the relative overlap of the intervals. Let the relative overlap between case i and case j, $\hat{\theta}_{ij}$be defined as

$$\hat{\theta}_{ij} = \frac{\theta_{ij}}{R_i}$$

Where θ_{ij} is the actual overlap between the two bootstrap intervals, and R_i is the range of the ith bootstrap interval. This way, the relative overlap is always 1 when $i = j$, and is closest to 1 two bootstrap intervals have the most overlap. This also allows for asymmetrical similarity. This way $\hat{\theta}_{ij} \neq \hat{\theta}_{ji}$ if the bootstrap intervals are not equally long. This metric was developed in order to avoid the problems that occur if one case has much higher variance than the others (as in Figure 3).

This similarity matrix was also used to implement the Bayesian X-means clustering method (Pelleg, Moore, 2000). X-means is an algorithm that increases the effectiveness of K-means clustering by optimizing the cluster locations and cluster space using either the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC) (Montgomery, Peck, Vining, 2015). Between the two clustering methods, AP-clustering proved to be more reliable with the dataset used within BRAWLER and CFAM. In instances where many cases had the same mean low variance, X-means either provided the same groups as AP-clustering or grouped those cases with other clusters, increasing the overall variance for each group.

4 IMPLEMENTATION AND FURTHER STUDY

Both the variance-grouping and AP-clustering methods show promise in grouping similar cases together. Unfortunately, due to the complex nature of the cases and scenarios utilized in BRAWLER and CFAM, no single method will always prove to be the best technique. One method for determining the validity of the groups is to compare the confidence interval for a given response with the sensitivity analysis of the mixed-integer program. Once CFAM has concluded its run, sensitivity analysis provides the acceptable ranges for each coefficient. If these coefficient ranges fall within the confidence interval for that response, it would suggest that the solution for CFAM is robust enough to account for the variance from BRAWLER. If the coefficient range is not within the confidence interval, additional clustering and/or comparison methods may need to be tried.

5 CONCLUSION

Integrating data from multiple modeling sources will always have complications. As models increase in scope and decrease in resolution, information and fidelity is lost in order to gain insights into the effects of larger moving parts. In the case of CFAM and BRAWLER, information about the nature of a given case's performance is lost when point estimates are used as inputs for the mixed-integer program. To account for this, a process has been put forward that is capable of conducting many comparisons of distributional data as well as three options for grouping cases based on the results of those comparisons. Using processes like this will hopefully improve model fidelity in other areas where stochastic models are used to provide inputs for deterministic models at higher levels. In the case of BRAWLER and CFAM, this process will improve AF/A9's ability to provide high-fidelity analysis to decision makers.

REFERENCES

Frey, B.J. and Dueck, D., 2007. "Clustering by passing messages between data points." Science 315(5814), : 972-976.

Montgomery, D.C., 2008. *Design and Analysis of Experiments.* John Wiley & Sons.

Montgomery, D.C., Peck, E.A. and Vining, G.G., 2015. *Introduction to Linear Regression Analysis.* John Wiley & Sons.

Mooney, C.Z., Duval, R.D. and Duval, R., 1993. "Bootstrapping: A nonparametric approach to statistical inference" (No. 94-95). Sage.

Pelleg, D. and Moore, A.W., 2000, June. "X-means: Extending K-means with Efficient Estimation of the Number of Clusters." In ICML (Vol. 1).

AUTHOR BIOGRAPHIES

BENJAMIN R. MAYO obtained an M.S. in Operations Research from the Air Force Institute of Technology in 2015. He has been a force structure mission analyst since 2015. His email address is benjamin.r.mayo.mil@mail.mil

TODD J. PACIENCIA obtained a PhD in Operations Research from the Air Force Institute of Technology in 2014. He has been a force structure campaign analyst since 2014. His email address is todd.j.paciencia.mil@mail.mil

DANIEL P. CROGHAN obtained a B.S. in Aerospace Engineering from the University of Maryland, and an M.S. in Operations Research from George Washington University. He has been an air to air combat analyst in the aerospace industry and the US Air Force since 1991. His email address is daniel.p.croghan.civ@mail.mil