

Determining A Minimal Alternatives Replication Set For Constructive Combat Simulation

Paul J. Deason, PhD
U.S. Army Conference on Applied Statistics
Georgia Tech, Atlanta
Oct 2004

Problem

- **CASTFOREM is a closed-form stochastic physics-effects based combat simulation.**
- **As more fidelity and capability to represent combat is added the single replication run time has become very large.**
- **A study consists of a Base Case and alternatives which are variations in weapons systems, etc.**
- **A usual study has 11 to 21 replications of the Base Case and each alternative**
- **A first-pass method to indicate where differences might reside using fewer replications is desired.**
- **Others have recommended the Boot Strap procedure for CASTFOREM. This is another idea.**

Assumptions

- **Measures of merit are**
 - **Force Exchange Ratio [(Red Loss/Initial Red)/(Blue Loss/Initial Blue)]**
 - **Blue Loss.**
- **A first-pass method is wanted to determine areas of potential differences, and where further exploration may be warranted.**
- **Traditional hypothesis testing is not feasible due to the number of replications required for precision, and the cost in obtaining them.**
- **The stochastic combat simulation running one scenario is a closed system.**
 - **A “run number” identifies a closed set of simulation “seeds” (see next slide)**
- **A finite number of replication runs of the base case stochastic simulation defines the population of interest.**

The Nature of CASTFOREM's Case Number Run Seeds

- In CASTFOREM, each replication run number identifies a set of number seeds used in the stochastic simulation.
- If there is no change in the model, scenario or data, and the same run number (with its underlying seeds) is used in a replication, results will be identical.
- Rather than using a random set of run numbers for producing comparison results, use a set centered on the Base Case central value
 - The run numbers responsible for the replication run(s) producing results close to the central measures of the Base Case for the principle measure of interest can be identified.
 - In the example, the 9th case has the median FER, and the 7th case the median Blue Loss.
- Seeds identified by these case numbers can then be used as seeds to run the alternate cases.
 - For example, sort in ascending order the FER and the associated run numbers
 - Select the run number for the Median, and the three adjacent runs below and above for a total of seven run seeds
- The objective would be for the bulk of the limited resources available for the simulation effort to be expended in running the Base Case.
 - A limited set of runs would be made for the alternatives based on the run numbers identified by the Base Case's Median and adjacent runs.

Why Select a Sample Based on the Base Case Center

- **The rationale behind selecting a sample determined by the central performance of the Base Case is to select a set that has the best possibility of exhibiting the same behavior.**
 - **Variance within the alternative sample is not of interest**
- **IF the limited sample is different enough to be identified as showing itself to be different in the most global measure of effectiveness (the most robust and resistant to change,) then there may be a strong likelihood that the alternative may in fact exhibit differences.**
- **Additional runs may then be undertaken to better understand the nature of the differences due to the alternative.**

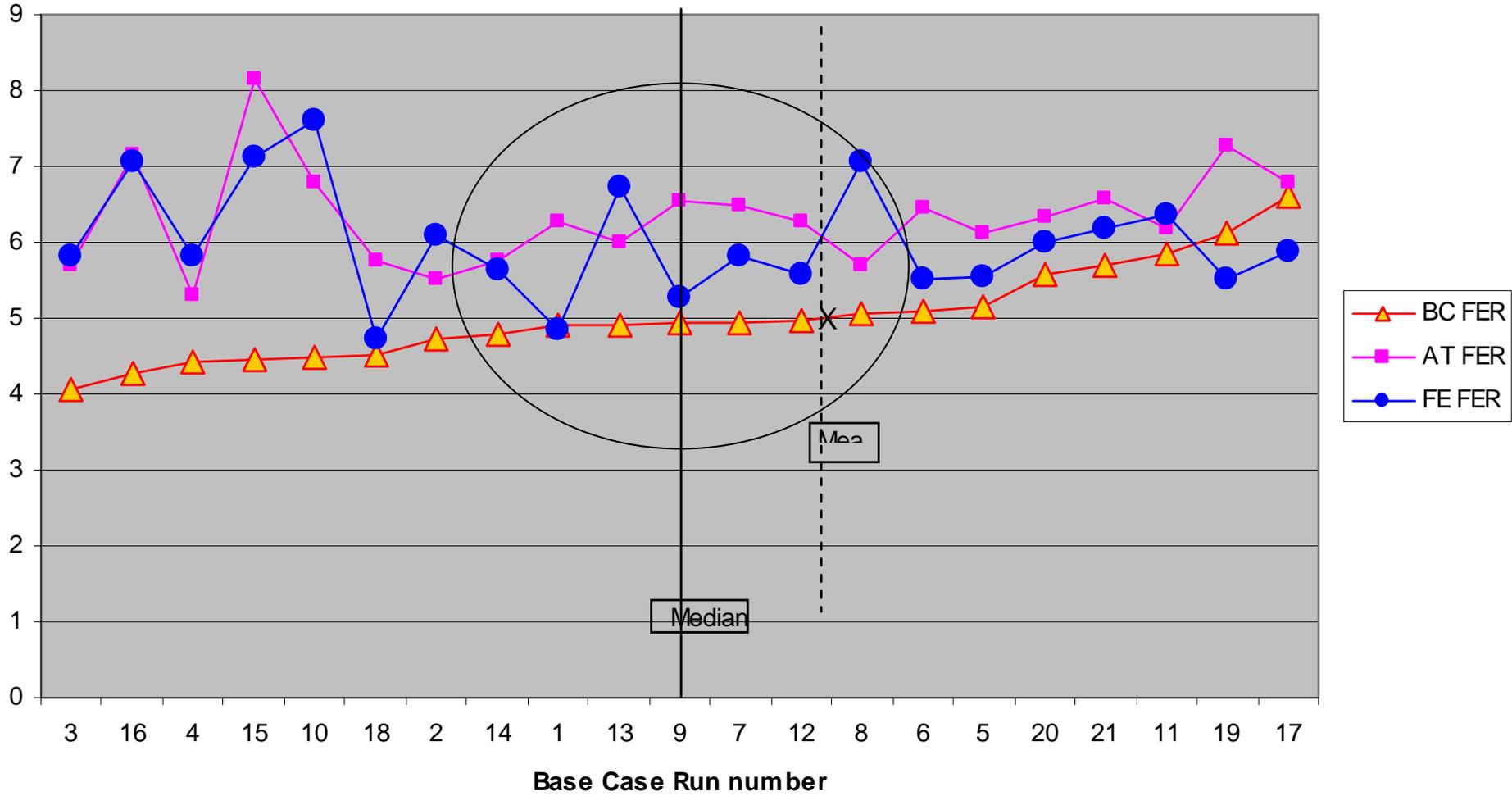
CASTFOREM Example

- **This is an unclassified example based on an Infantry Battalion in a fixed point defense against a Threat armored brigade**
 - The Base Case had the usual direct fire and artillery support
 - One alternative had a unit of advanced anti-armor (AT) weapon system
 - One alternative had a system of systems for defeating Armor deep, with multiple entities, interlinked systems of anti-tank, artillery, and other combat multipliers. (This simulation scenario was based on a field experiment)
- **Principle measure of effectiveness-- Measures of merit**
 - Force Effectiveness Ratio $[(\text{Red Loss}/\text{Initial Red})/(\text{Blue Loss}/\text{Initial Blue})]$
 - Blue Loss (interest actually in Blue survival)
 - These are non-independent measures

Method using Median

- **Run Base Case numerous times (21 in this case)**
- **Select median 7 replications (median, three up, three down) based on:**
 - **Force Exchange Ratio, or**
 - **Blue Loss (they are not independent)**
- **Run the alternative using the 7 run “seeds” identified by their run numbers in the BC**
- **Test by:**
 - **Parametric (ANOVA, SNK, Isd)**
 - **Non-parametric (Kruskal-Wallis, Median test, Mann-Whitney)**

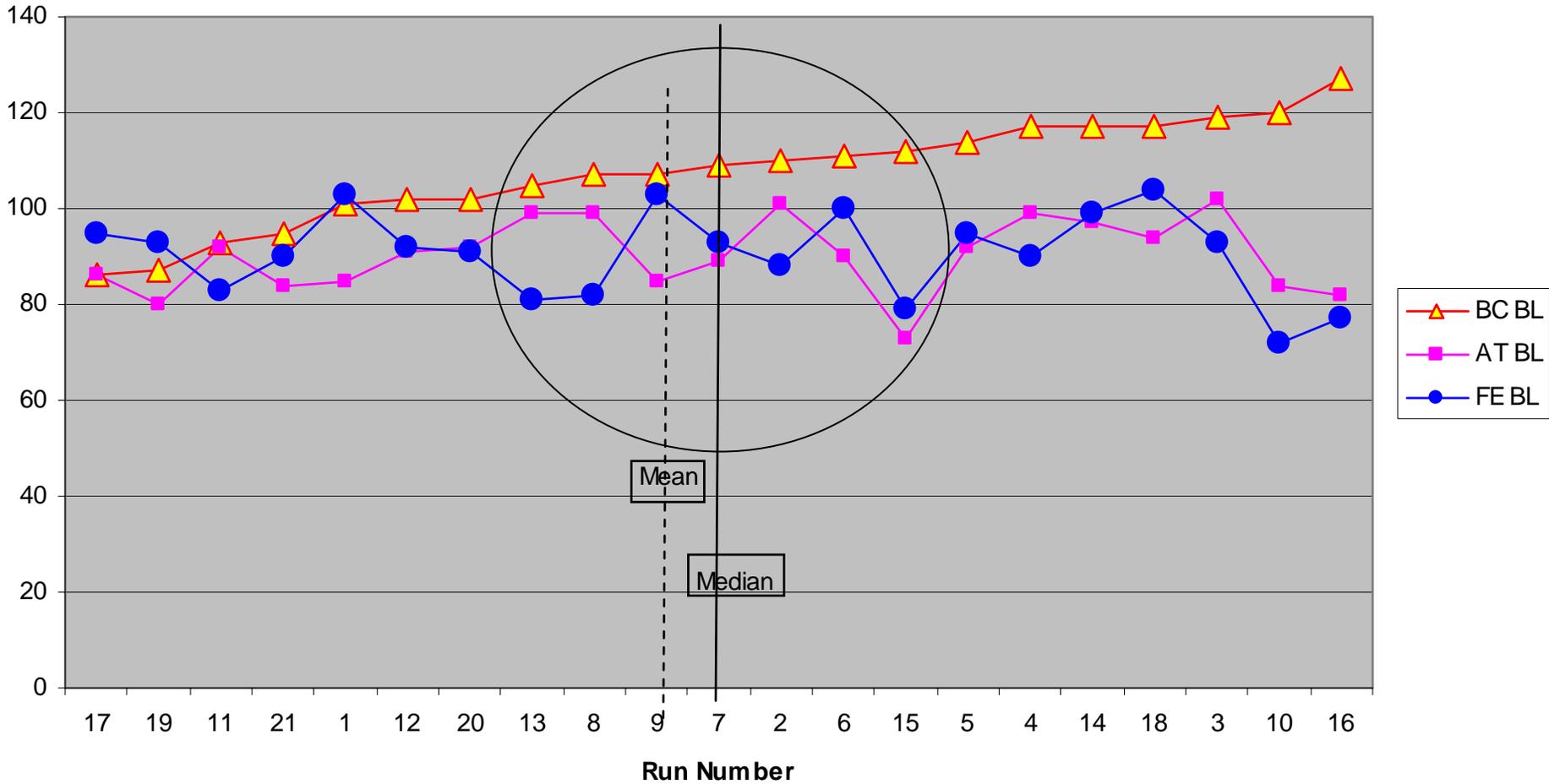
FERs Ordered by Base Case FER



Statistical Tests Conducted on FER

- **Full sample of 21 replications per case**
 - **Variance homogeneous** (Levene (2,60)=.414 p=.663)
 - **One-Way ANOVA significant <.001;**
 - **SNK, Isd => BC<AT=FE**
 - **Kruskal-Wallace significant <.001; Median test < .001**
 - **Pairwise Mann-Whitney => BC<AT; BC<FE; AT=FE**
- **Reduced – 21 for Base Case, 7 for alts**
 - **Variance homogeneous** (Levene (2,32)=.1.221 p=..308)
 - **One-Way ANOVA significant <.001, Isd=> BC<AT=FE**
 - **Kruskal-Wallace significant =.001; Median test < .001**
 - **Pairwise Mann-Whitney => BC<AT; BC<FE; AT=FE**

Blue Loss ordered by Base Case



Statistical Tests Conducted on Blue Loss

- **Full sample of 21 replications per case**
 - **Variance homogeneous** (Levene (2,60)=1.071 p=.349)
 - **One-Way ANOVA significant <.001;**
 - **Isd, SNK => BC>AT=FE**
 - **Kruskal-Wallace significant <.001; Median test < .001**
 - **Pairwise Mann-Whitney => BC>AT; BC>FE; AT=FE**
- **Reduced – 21 for Base Case, 7 for alts**
 - **Variance homogeneous** (Levene (2,32)=.0094 p=..910 (suspicious?))
 - **One-Way ANOVA significant <.001;**
 - **SNK, Isd=> BC>AT=FE**
 - **Kruskal-Wallace significant =.001; Median test < .001**
 - **Pairwise Mann-Whitney => BC>AT; BC>FE; AT=FE**

Another Approach? – Method using Mean

- **Take mean, SE, n of the Base Case sample**
- **Sort the Base Case results in ascending order & select 7 replications**
 - **the one nearest the mean, three above and three below the mean value and note their case numbers.**
- **Run the 7 cases in the alternate using the same case number “seeds”**
- **Do a t-test between the mean value of the Base Case and the mean of the alternate sample. Use the SE $[sd/(n)^{1/2}]$ of the Base Case for the comparison.**

Ideas?

- **This presentation is in the pure sense of a clinical case.**
- **Using the Median or the Mean as a marker in CASTFOREM for replications is an idea that seems plausible.**
- **Is this approach logical?**
 - We evaluate at a system of systems level, not one measure.
 - Usually the first measure we are interested in is mission accomplishment, the others are side issues. So for different measures would this reduced population have different base populations, would we end of running more runs?
- **Is another approach better when the intent is a limited number of costly replications to identify potential differences in alternatives?**
 - For instance, the Median/hinge Gaussian comparison (Velleman & Hoaglin)
 - Cioppa's Boot Strap?

Paul J. Deason, Ph.D.

paul.deason@us.army.mil

505-678-1610 (DSN 258)