Preference-based Monte Carlo Weight Assignment for Multiple-criteria Decision Making in Defense Planning*

Johan Schubert, Pontus Hörling
Department of Decision Support Systems
Division of Information and Aeronautical Systems
Swedish Defence Research Agency
SE-164 90 Stockholm, Sweden
schubert@foi.se
http://www.foi.se/fusion

***Abstract*—In this paper we develop a multiple process approach for finding the most preferred plan instance within military defense planning. In this problem tens of thousands of alternative plans with different simulation settings are evaluated by multiple measures of effectiveness through simulation. In a sequence of processes, the Pareto optimal frontier of the plans is first found. Using decision makers' preferences where the decision makers may provide incomplete preferences regarding subsets of disjoint measures, these measures are ordered by their importance. Alternative sets of random weights are assigned to the set of measures by a Monte Carlo approach where weights uphold the preferred order. Finally, all plans are evaluated using the weighted measures and the most preferred plans according to a ranking index are further analyzed to provide decision support.**

Keywords—Defense planning; Pareto optimal frontier; preference modelling; belief function; Dempster-Shafer theory; multiple-criteria decision making; MCDM; Monte Carlo; ranking; data analysis.

I. Introduction

In this paper we develop a multi-criteria decision support methodology for defense analysis. With this methodology a decision maker can analyze tens of thousands of alternative plan instances regarding the best use of available resources in military operations. The analysis use data that is generated by a simulation system with stochastic processes that simulates alternative parameter settings for the scenario under analysis. In an experiment we use a ground warfare scenario where 50 000 alternative plan instances described by 37 parameters each, are simulated 20 times with different random seeds. Each simulation is evaluated by multiple measures of effectiveness (MOE) that measure how well each individual scenario instance performs in fulfilling all of the criteria.

Initially, we try to find the most effective plan instance as evaluated by all measures of effectiveness (MOEs). When there are many such measures we are faced with a multiple-criteria decision making problem when assessing which plan instances are preferred. As a first step, we develop a new method for finding the Pareto optimal frontier [1–2] of the entire set of all plan instances where utility intervals over each measure of effectiveness is received from the 20 simulations, due to the stochastic nature of the simulations. The plans on the Pareto optimal frontier are the plans that are better than all other plans regardless of how the measures of effectiveness might be weighted in a subsequent assessment process.

As we assume that it is impossible for decision makers to assign precise weights to all measures of effectiveness, we let a group of decision makers express any number of preferences on the importance between any two disjoint subsets of MOEs. We further develop an extension to Utkin's [3] preference ranking method which is focused on finding the order of importance of the measures of effectiveness from the preference assignments made by the decision makers. Utkin's method is extended by Schubert's method [4] for interpolation in belief-plausibility intervals regarding the obtained degree of preference of all different measures of effectiveness.

Using the preference order of importance for all MOEs we develop a Monte Carlo approach for assigning weights to these MOEs. In this method we randomly assign weights that abide by the preferred order of importance of the MOEs. That is, the most preferred measure will be weighted higher than the second most preferred measure, etc. In the spirit of the Monte Carlo approach we perform 1000 alternative weight assignments for all measures, yielding 1000 alternative rankings of all alternative plan instances. The plans with the highest average ranking over the 1000 alternative rankings are the most preferred plan instances. A process overview is provided in Fig. 1.

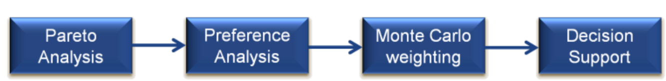

Fig. 1. Process overview.

Other authors have considered different approaches to weight assignment. Huang *et al*. [5] consider the assignment of weights to criteria based on the consistency and similarity of the opinions from decision makers regarding these criteria. In addition it is also possible to let the decision makers themselves be weighted. Yue [6] suggest using the decision makers' experience regarding the topic under consideration as

*This work was supported by the FOI research project "Simulation-based Defense Planning", which is funded by the R&D programme of the Swedish Armed Forces.

a basis for assigning weights. A third approach discussed in this paper, is to let each decision maker use a weighting of his own as an expression of the importance placed on a pairwise comparison of two disjoint subsets of criteria.

In Sec. II we describe defense planning as a concept. In Sec. III we describe our multi-criteria approach to decision making. In Sec. IV we develop an approach for finding the Pareto optimal frontier for a multiple run stochastic data farming experiment with interval parameter evaluation. We then develop the extension to Utkin's preference ranking method (Sec. V). In Sec. VI we use the Monte Carlo approach for finding 1000 alternative random assignments of weights to all measures of effectiveness. In Sec. VII we analyze the ranking order for all 1000 alternative rankings, and find the best and worst ranking index for all remaining plan instances over all 1000 alternative rankings as a final meta-ranking. This meta-ranking is used to find the best plans that are further analyzed in order to explain the reason for success of the best plans. Finally, conclusions are drawn (Sec. VIII).

II. Defense Planning

In defense planning the overall question is how to best use all available resources in military operations. Some examples of more detailed questions under study are:

- How do we manage our resources for maximum effectiveness?
- What are the requirements on sensors, weapons, tactics, etc. to achieve success?
- Which overall configuration of forces is most effective?

To answer these questions we simulate tens of thousands of scenario instances and perform extensive data analysis to find the scenario instances with the most promising parameter setting. For these particular instances we investigate which parameters and what combination of parameter ranges lead to overall success.

A. Scenario

The scenario of study in this paper is ground warfare between two opposing battalions. On the blue defending side we model a mechanized battalion with 60 *Combat Vehicle 90*. On the red attacking side we model a parachute battalion with 27 *Airborne Combat Vehicle BMD-4* and 16 *Armored Personnel Carrier BTR-D*. In the scenario the red force is on the move attacking the defending blue force, with an objective to break through.

In order to learn as much as possible from the simulations as a data farming experiment, we will vary the numbers of units on both sides in different simulation runs, as well as study the effects of different avenues of approach, different tactics, different sensor ranges, etc.

III. Multi-criteria Decision Making

When simulating the ground warfare scenario we evaluate the outcome by 37 different input parameters. These are 32 binary (Boolean) parameters and five real parameters. Of these, 18 parameters describe input values for the blue side, and 19 parameters describe the red side. In the simulations we vary all of these input parameters for 50 000 different plan instances P_i, $1 \leq i \leq 50\,000$.

We measure the effect by five measures of effectiveness. The *utility value* of MOE number j in plan P_i is designated as E_{ij}. Each of the plans is simulated 20 times with the same set-up of input parameter values, only with different seeds. In total we have 1 000 000 simulated outcomes of all plans as each P_i results in a group of 20 outcomes. In P_i we designate the utility value of the k^{th} outcome for measure E_{ij} as E_{ijk}, where $1 \leq i \leq 50\,000, 1 \leq j \leq 5, 1 \leq k \leq 20$. In total we have 5 000 000 output values.

IV. Pareto Analysis

After simulating the 50 000 plans 20 times each with 20 different seeds, a first filtering of the output data set is made using Pareto analysis. This is a form of multi-criteria analysis. The simulation results are analyzed for the most important input and output parameters found. In the analysis we focus on two input parameters and three output measures of effectiveness;

- input: "# Blue Platoons In scenario" (*BPI*); it is a superior achievement to achieve success with a small number of blue initial platoons, than with a larger force,
- output: "# Blue Unit Losses" (*BUL*); we want to minimize blue losses as much as possible,
- input: "# Red Platoons In scenario" (*RPI*); it is a superior achievement to achieve success facing a large number of red initial platoons, than with a smaller red force,
- output: "# Red Unit Losses" (*RUL*); we want to maximize red losses as much as possible,
- output: "# Red Units Finish" successfully (*RUF*); we want to allow as few as possible of the enemy to break through the blue line of defense.

These five criteria can be simultaneously maximized or minimized, where the 1st, 2nd and 5th are minimized and 3rd and 4th maximized. These criteria can be weighted by their relative degree of importance. However, this is not needed in the Pareto analysis, which filters away all plan instances with a worse value on *all* criteria compared to *at least one other* plan instance among the set of 1 000 000. Some words are needed on how plans are compared in our Pareto analysis. First, we multiply the three parameters that are to be minimized by -1. Thus, criteria fulfillment, for all parameters will be better with higher parameter values. Since each plan is simulated 20 times with different seed, we get a variation of the three output measures (*BUL*, *RUL* and *RUF*) within a group of 20 seeds. When comparing two plans, we compare the *min–max* intervals for these three output measures. If there is an interval overlap for at least one of these three measures, no filtering will take place as a result of this comparison. For the two input parameters used in filtering (*BPI*, *RPI*) they are the same within each 20-seed group and comparison is made point-wise between plans. Hence, a plan P_l in a planning problem is filtered away if

$$\exists i \forall m.\ \min_k E_{imk} > \max_k E_{lmk}. \tag{1}$$

The idea is illustrated in Fig. 2. In our example, Pareto filtering 50 000 plans leave 31 920 of them on the Pareto optimal frontier. Note here, that even though we had a total of 37 input parameters and five output measures characterizing each plan instance that we could impose criteria on, we chose to use only the five most discriminating parameters as MOEs in the Pareto analysis.

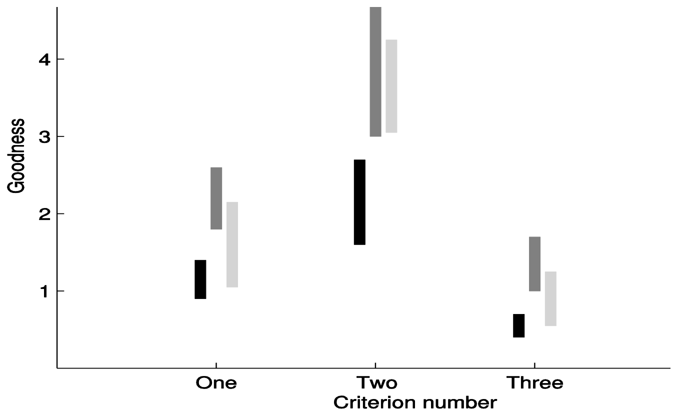

Fig. 2. A simple example with three plans (black, dark grey, light grey), and how their simulated outcomes from different seeds (giving values within a bar, indicating a *min–max* interval) fulfill three parameter criteria intervals. The two grey plans are seen to be Pareto optimal. The black plan can be filtered away after comparison with the dark grey plan.

This approach is independent on any individual criteria weighting. After comparing all plans to each other, the remaining plans are now said to be on the Pareto optimal frontier. For these plans, not all criteria outcomes are worse than in some other plan.

Another kind of weighting using regression trees can be done in order to figure out which input parameters, and what ranges, most strongly affect certain output parameters. From a regression tree that is "trained" on a set of plan instances, we can roughly predict an output parameter value when we know the values of the input parameters for some new plan that is simulated with the same simulation processes that produced the training plan set. We use all input parameters for the regression analysis and the function `RegressionTree.fit` in MATLAB Statistical Toolbox. As an example, for non-Pareto-filtered data, we can check how the output parameter red units that crossed the finish line (*RUF*) can be estimated from input parameters. This is shown in Fig. 3.

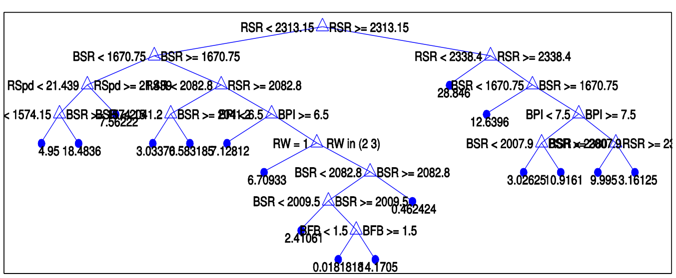

Fig. 3. A regression tree on the full plan set without any Pareto-filtering. Traversing the tree from root to leaf, at every branching point we select the branch with the value we have for the given input parameter at that branching point. Eventually, we reach a leaf which predicts the resulting output parameter value. Red sensor ranges are in top, giving highest impact on *RUF*.

When working on the smaller Pareto-filtered set, we get the regression tree in Fig. 4. Now, the initial number of blue platoons (*BPI*) is the by far most important parameter.

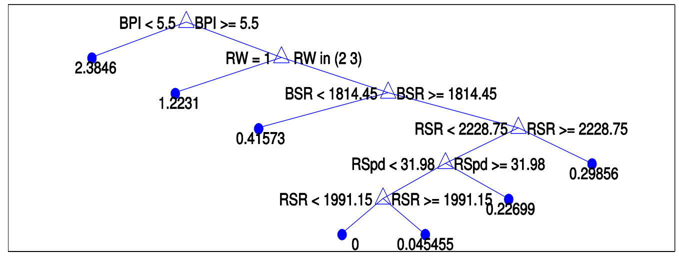

Fig. 4. A regression tree on the Pareto-filtered plan set. Here size of own forces (*BPI*) have strongest impact on *RUF*.

We can also depict the total relative importance of the input parameters on an output parameter. We can compute estimates of input parameter importance by summing changes in mean squared errors due to splits in the regression tree on every parameter, and dividing the sum by the number of tree nodes. For the non-Pareto-filtered case, we observe in the upper part of Fig. 5 that *RUF* is mostly impacted by blue and red sensor ranges (*BSR*, *RSR*), which is reflected by the uppermost branching points in the regression tree. This is the case when working on the non-filtered plan set of consisting of all plans, i.e., also including not so successful plans. When working on the Pareto-filtered set (lower part) we notice the importance of *BPI* and *RW*.

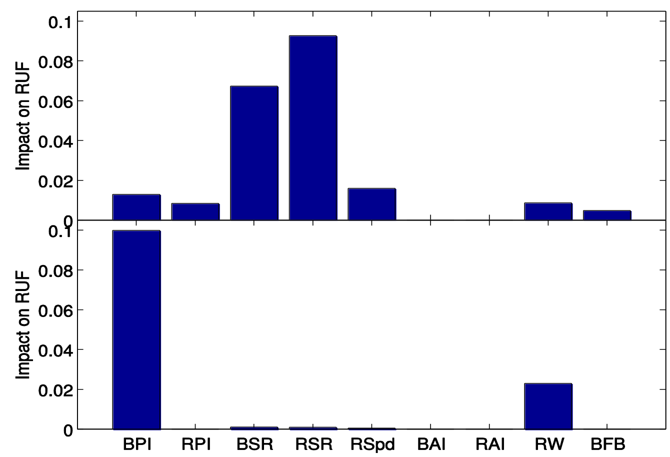

Fig. 5. The importance of some input parameters on *RUF* for the full non-Pareto-filtered plan set (*upper*) and the Pareto-filtered plan set (*lower)*. The difference is large indicating that, for the best plans, the size of own forces (*BPI*) is of very large importance for the amount of red units that succeeds in crossing the finish line (which we want to prevent). When including plans of lower quality, sensor ranges tend to much more affect the outcome. In neither case the initial amount of ammunition (*BAI*, *RAI*) affects the outcome as neither of the parties consumes all their ammunition.

V. DECISION MAKERS' PREFERENCES

In order to be able to rank all simulations based on their measure of effectiveness (MOE) we need to be able to assign weights to these measures. This is something that is usually difficult for a decision maker. On the other hand, it is often possible to express an order of importance between different measures, or at least between some subsets of measures.

In this section we will derive a method for finding an exact preference order of all measures of effectiveness. This method will accept any preference expression about the MOEs from multiple decision makers. For example, expressions such as *"measure of effectiveness number i is more important than measure of effectiveness number j"*; $MOE_i \succcurlyeq MOE_j$, or expressions regarding two different subsets of all measures such as *"measures i and j are more important than measures k and l"*; $\{MOE_i, MOE_j\} \succcurlyeq \{MOE_k, MOE_l\}$. We use a preference assignment approach developed by Utkin [3] in combination with a preference ranking approach by Schubert [4] to derive a complete ranking of all MOEs.

We will keep track of all preferences expressed by all decision makers. This includes both preferences about the order of importance among single measures and among subsets of measures. For each expression we count the number of decision makers giving the same preferences. (Alternatively, we could let a decision maker use a weighting of his own expression of importance, where each decision maker can increase the count for a preference by more than one; this is, however, not used in this paper).

For each preference we sum the total number of assigned preferences by all decision makers

$$c_{AB}\left(\{MOE_i\}_{i\in A} \succcurlyeq \{MOE_j\}_{j\in B}\right), \quad (2)$$

where $\emptyset \neq A, B \subseteq \{i\}_{i=1}^{|\Theta|} = I$, i.e., A and B are subsets of an index set I of indices corresponding to the set of all MOEs, $\Theta = \{MOE_i\}_i$. Any number of these c_{AB} may be equal to zero, due to a lack of assigned preferences regarding some subsets of MOEs.

The preferences assigned between *subsets* of measures can be simplified to a set of preferences among *single* measures [3]. We have,

$$\{MOE_i\}_{i\in A} \succcurlyeq \{MOE_j\}_{j\in B} = \{MOE_i \succcurlyeq MOE_j\}_{i\in A, j\in B}. \quad (3)$$

From the counts of assigned preferences in (2) we derive a basic belief assignment within belief function theory [7–9]. In this setting of our problem representation, the frame of discernment (i.e., the set of all possible elementary preferences) is

$$\Omega = 2^{\{MOE_i \succcurlyeq MOE_j\}_{i,j\in I}}. \quad (4)$$

We have the following basic belief assignment,

$$m_{AB}\left(\{MOE_i \succcurlyeq MOE_j\}_{i\in A, j\in B}\right) = \frac{1}{N} c_{AB}\left(\{MOE_i \succcurlyeq MOE_j\}_{i\in A, j\in B}\right) \quad (5)$$

where N is the total sum of all counts

$$N = \sum_{AB} c_{AB}\left(\{MOE_i \succcurlyeq MOE_j\}_{i\in A, j\in B}\right). \quad (6)$$

While it is possible to change the representation in (5) using (3), it is not possible to divide the basic belief mass among the different preferences in $\{MOE_i \succcurlyeq MOE_j\}_{i\in A, j\in B}$ as we have no information on how to divide it among the different preferences. Instead the entire mass must remain on the whole set.

From the basic belief assignments in (5) we can calculate belief and plausibility for any subset of the frame of discernment. However, our interest is limited to analyze the support received by all preferences regarding single measures of performance such as,

$$\{MOE_i\} \succcurlyeq \Theta, \forall i \quad (7)$$

where $|\{MOE_i\}| = 1$. These preferences are what we need, to make a full preference based ranking of all MOEs.

We have belief of the preference of MOEs,

$$Bel_{\{i\}\Theta}(\{MOE_i\} \succcurlyeq \Theta) = Bel_{\{i\}\Theta}\left(\{MOE_i \succcurlyeq MOE_j\}_{j\in I}\right) = \sum_{X \subseteq \{MOE_i \succcurlyeq MOE_j\}_{j\in I}} m_{\{i\}B}(X) \quad (8)$$

and plausibility

$$Pls_{\{i\}\Theta}(\{MOE_i\} \succcurlyeq \Theta) = Pls_{\{i\}\Theta}\left(\{MOE_i \succcurlyeq MOE_j\}_{j\in I}\right) = \sum_{X \cap \{MOE_i \succcurlyeq MOE_j\}_{j\in I} \neq \emptyset} m_{\{i\}B}(X). \quad (9)$$

We can sort all preferences based on the belief and plausibility for each preference $\{MOE_i\} \succcurlyeq \Theta$.

When both

$$Bel_{\{i\}\Theta}(\{MOE_i\} \succcurlyeq \Theta) > Bel_{\{j\}\Theta}\left(\{MOE_j\} \succcurlyeq \Theta\right) \quad (10)$$

and

$$Pls_{\{i\}\Theta}(\{MOE_i\} \succcurlyeq \Theta) > Pls_{\{j\}\Theta}\left(\{MOE_j\} \succcurlyeq \Theta\right) \quad (11)$$

then $MOE_i \succcurlyeq MOE_j$.

When an interval $\left[Bel_{\{j\}\Theta}, Pls_{\{j\}\Theta}\right]$ is included in an interval $\left[Bel_{\{i\}\Theta}, Pls_{\{i\}\Theta}\right]$ it is not immediately clear which is the preferred measure; MOE_i or MOE_j. We can interpolate with a parameter $\rho \in [0, 1]$ in each belief-plausibility interval in order to find the preferred measure [4]. However, we have no information regarding the value of ρ, and any assumption about ρ will be unwarranted.

Instead we may calculate the point ρ_{ij} where the two measures MOE_i and MOE_j are equally preferred. When

$$\left[Bel_{\{i\}\Theta}, Pls_{\{i\}\Theta}\right] \supset \left[Bel_{\{j\}\Theta}, Pls_{\{j\}\Theta}\right] \quad (12)$$

correction inserted

we have

$$\rho_{ij} = \frac{Bel_{\{j\}\Theta} - Bel_{\{i\}\Theta}}{\left(Pls_{\{i\}\Theta} - Bel_{\{i\}\Theta}\right) - \left(Pls_{\{j\}\Theta} - Bel_{\{j\}\Theta}\right)} \quad (13)$$

where each belief and plausibility function is taken for $\{MOE_i\} \succcurlyeq \Theta$ and $\{MOE_j\} \succcurlyeq \Theta$, respectively. If $\rho_{ij} < 0.5$ then $MOE_i \succcurlyeq MOE_j$.

When these points of equal preferences are calculated for all points ρ_{ij} we can order all measures in a chain of preference. The different MOEs will be preferable for different intervals of ρ. Some MOEs may sometimes be dominated by other MOEs and not be preferable for any interval length of ρ. However, this is only important when searching for the MOE with highest probability of being the most highly preferred MOE among the set of all MOEs; $\{MOE_i\}_i$ [4].

We notice that in the special case when we are only comparing MOEs pairwise one-by-one the situation is simplified. The requirement that we must have $\rho_{ij} < 0.5$ in order for $MOE_i \succcurlyeq MOE_j$ is equivalent to having

$$Bel_{\{i\}\Theta} + \frac{1}{2}\left(Pls_{\{i\}\Theta} - Bel_{\{i\}\Theta}\right) > Bel_{\{j\}\Theta} + \frac{1}{2}\left(Pls_{\{j\}\Theta} - Bel_{\{j\}\Theta}\right), \quad (14)$$

i.e., that the mid-point in the belief-plausibility interval MOE_i is higher than for MOE_j.

This implies that we can obtain an exact preference order of all MOEs using a standard sorting algorithm based on the belief-plausibility interval mid-points for each MOE.

A. An example

Based on a pre-evaluation of the importance of different measures, a set of five measures of effectiveness was deemed important for assessment of different simulation input parameter instances. We use $\{MOE_i\}_{i=1}^{5}$ in this example, where

- MOE_1 is "# Blue Platoons In scenario" (*BPI*),
- MOE_2 is "# Blue Unit Losses" (*BUL*),
- MOE_3 is "# Red Platoons In scenario" (*RPI*),
- MOE_4 is "# Red Unit Losses" (*RUL*),
- MOE_5 is "# Red Units Finish successfully" (*RUF*),

and MOE_1 and MOE_3 play a dual role as input parameters that are also used as measures of effectiveness.

In this example we have 76 statements of preference order. These are preference given by several experts. In contrast to the approach taken by Utkin [3] we register all given preference equally regardless of whether they are given by one or several decision makers.

In Table I we observe a few of the preferences given by multiple experts and the recorded number of experts for that preference.

TABLE I. NINE SELECTED PREFERENCES AS EXAMPLES OUT OF 76.

PREFERED SET	$\succcurlyeq$	NON-PREFERED SET	C_{AB}
{*BUL*}	$\succcurlyeq$	{*BPI, RPI, RUL*}	5
{*RPI*}	$\succcurlyeq$	{*BPI*}	2
{*RUL*}	$\succcurlyeq$	{*BPI, RPI*}	3
{*RUF*}	$\succcurlyeq$	{*BPI, RPI, RUL*}	1
{*BPI, BUL*}	$\succcurlyeq$	{*RPI, RUL*}	1
{*BUL, RUL*}	$\succcurlyeq$	{*BPI, RPI*}	1
{*BUL, RUF*}	$\succcurlyeq$	{*BPI, RPI*}	1
{*RUL, RUF*}	$\succcurlyeq$	{*BPI, BUL, RPI*}	1
{*BUL, RUL, RUF*}	$\succcurlyeq$	{*BPI, RPI*}	1

From the given preference we assign basic beliefs using (5) and calculate belief and plausibility for all $\{MOE_i\} \succcurlyeq \Theta$ using (8) and (9). These results are presented in Table II.

TABLE II. SUPPORT FOR MOES

	MOE_1 ***BPI***	MOE_2 ***BUL***	MOE_3 ***RPI***	MOE_4 ***RUL***	MOE_5 ***RUF***
Bel	0	0.4605	0.0263	0.1184	0.0921
Pls	0.0395	0.6711	0.0263	0.2895	0.3158

From these numbers it is immediately obvious that "# Blue Unit Losses" (*BUL*), is the most preferred MOE since $Bel(MOE_2) > Pls(MOE_i), \forall i \neq 2$. We also notice that the belief-plausibility interval of MOE_4 is included in that of MOE_5, and MOE_3 is included in MOE_1. As $Bel(MOE_5) > Pls(MOE_1)$ we may conclude that $\{MOE_4, MOE_5\} \succcurlyeq \{MOE_1, MOE_3\}$ but we don't know directly which of MOE_4 and MOE_5, on the one hand, and MOE_1 and MOE_3, on the other hand, that are the preferred measure. Using the ranking of (14) with midpoints $\rho_{13} = \rho_{45} = 0.5$ we get,

$$BUL \succcurlyeq RUL \succcurlyeq RUF \succcurlyeq RPI \succcurlyeq BPI, \quad (15)$$

as a full preference ordering of all measures of effectiveness.

This is the preference order (15) that we will use in the following sections.

VI. MONTE CARLO WEIGHTING

Weighting of the different criteria can be done if one has sufficiently good knowledge of which parameter inputs and simulation outputs that are most important to minimize or maximize in order to obtain a preferred outcome in a planning process. Often, only simple preferences can be given, like "MOE_2 is more important to maximize than MOE_4".

When one measure is preferred over another the weight of that measure must be higher than the other. With two measures with $MOE_1 \succcurlyeq MOE_2$ we may plot weight-pairs of these measures in a two dimensional diagram. If MOE_1 is on the *y*-axis and MOE_2 is on the *x*-axis all allowed weight-pairs will be in the upper left triangle. With preference orders for more measures, the allowed volume will be a cut-out of a hypercube.

Now that we have a full preference ordering (15) of all measures, we still have not assigned any weights (obeying this preference ordering) for further analysis. We reduce the problem step-by-step before assigning the weights: We have

$P_i, 1 \leq i \leq 31\,920$ remaining plans instances after the Pareto filtering of five criteria. We will replace the set of 20 outcomes for each measure from the 20 simulations of a particular plan instance P_i with a two point interval using the *min* and *max* values of from the set of 20, analogously to the Pareto analysis above, (designated with an asterisk "*" at the lower and upper index level) and assigning the resulting values to corresponding lower and upper interval points, where

$$P_{i*}: \ E_{ij*} = \min_k E_{ijk}, \tag{16}$$

$$P_i^*: \ E_{ij}^* = \max_k E_{ijk} \tag{17}$$

are framing each plan P_i.

We then normalize the values for all measures in such a way that for a chosen measure, its value equals zero if it has the lowest value among *all* Pareto-optimal plans, and equals one if it has the highest value among those plans. Hence, all measures will be distributed between 0 and 1. We have,

$$E_{i*} = \min_j E_{ij*}, \tag{18}$$

$$E_i^* = \max_j E_{ij}^* \tag{19}$$

and

$$E_{ij*}^N = \frac{E_{ij*} - E_i^*}{E_i^* - E_{i*}}, \tag{20}$$

$$E_{ij}^{N*} = \frac{E_{ij}^* - E_i^*}{E_i^* - E_{i*}}. \tag{21}$$

Our idea is to use a Monte Carlo approach to observe the robustness in ranking all 31 920 plan instances for different combinations of weights. Therefore, we sample 1000 ordered sets of five random weights $w_{js}, 0 \leq j \leq 5, 0 \leq s \leq 1000$ uniformly distributed $0 \leq w_{js} \leq 1$ where, within each set, the random numbers are sorted so they obey the preference ordering in (15). For each 1000 sets of random weights, we compute alternative upper and lower *utilities* R_i of all 31 920 plans. We have for each plan P_i,

$$R_{is*} = \frac{1}{5}\sum_{j=1}^{5} w_{js} \cdot E_{ij*}^N, \tag{22}$$

$$R_{is}^* = \frac{1}{5}\sum_{j=1}^{5} w_{js} \cdot E_{ij}^{N*}. \tag{23}$$

These pairs of upper and lower bounding for each plan can be seen as 1000 alternative expected utility intervals for each plan.

VII. Decision Support

In this section we focus on providing decision support regarding what is the best plan, and explain the reason for success of the best plan. We start with the just computed 1000 alternative utility intervals for each plan P_i. Each result is summed up over all alternative weight assignments and normalized to obtain a grand total *rank* for all plan instances bounding each plan,

$$R_{i*} = \frac{1}{1000}\sum_{s=1}^{1000} R_{is*}, \tag{24}$$

$$R_i^* = \frac{1}{1000}\sum_{s=1}^{1000} R_{is}^*. \tag{25}$$

The ranks can be visualized in different ways. To get a general view of the distribution of plan ranks we plot the five percentiles 0, 25, 50, 75, 100 corresponding to the *min, lower quartile, median, upper quartile* and *max* of all plan rankings R_i, see Fig. 6.

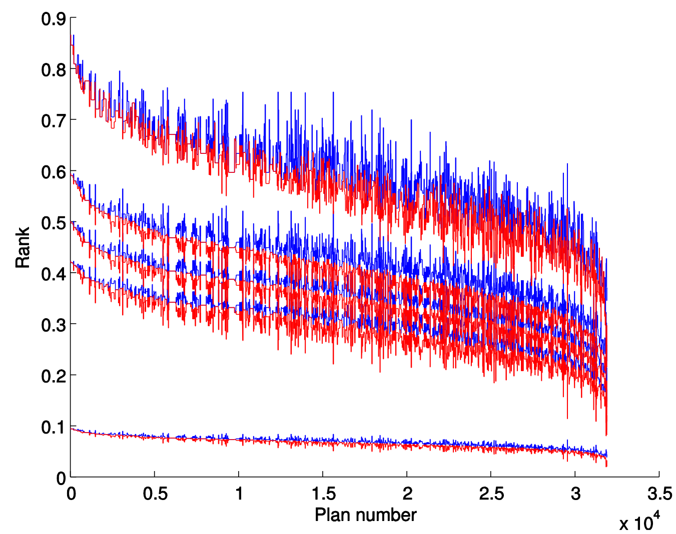

Fig. 6. The normalized ranks for all plans after Pareto-filtering. The five different percentiles are shown in red and blue for R_{i*} and R_i^*, respectively.

We also discriminate between R_{i*} and R_i^* to show the plan intervals. The plans are sorted according their *mean* ranks, see Fig. 7. Therefore, the plots look "noisy" when plotting the percentiles. This is done simply to get a hint on the spread.

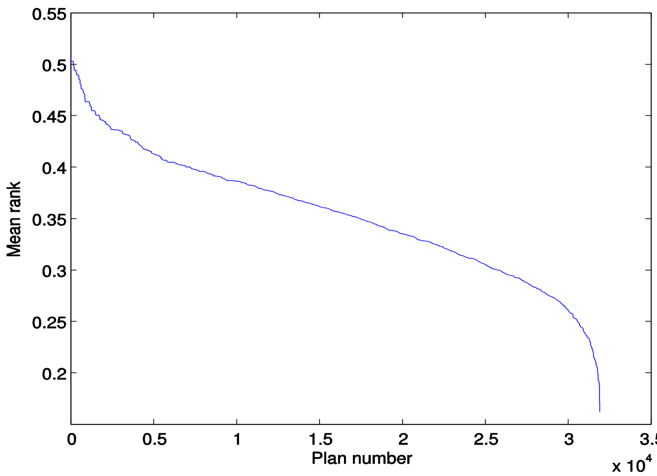

Fig. 7. The sorted mean normalized rank distribution of all Pareto-filtered plans.

In order to see how the plans are ranked for a more even planning situation, one can as an example use the subset where the number of blue and red platoons are equal to seven in both cases, that is *BPI* = *RPI* = 7. This constraint leaves 1500 plans out of the 31 920 plans to be analyzed. The rank distribution for those plans can be seen in Fig. 8. The plans are also here sorted according their *mean* ranks, see Fig. 9.

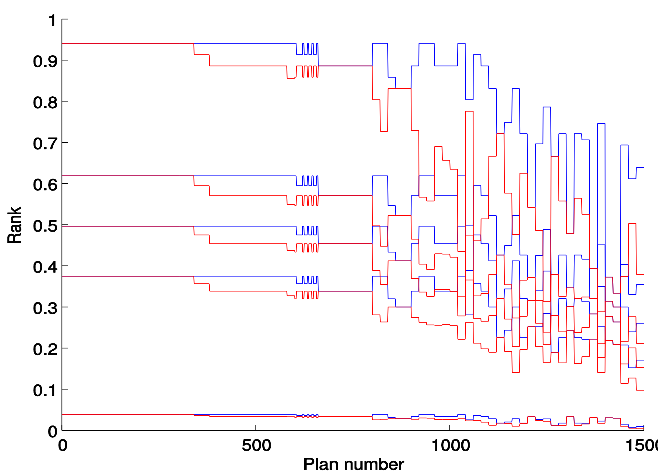

Fig. 8. The normalized ranks for all Pareto-filtered plans after selecting plans where seven blue platoons are set up against seven red platoons (*BPI* = *RPI* = 7). The five different percentiles are shown in red and blue for the then remaining R_{i*} and R_i^*, respectively.

We can see that the rank is grouped in levels or steps with the same rank value, for instance the highest rank values within each percentile are the same for the first approximately 350 plans, then it starts to drop step-wise. This first group of plans with approximately 350 plans is the solution to finding the best plans. The final step of analysis is to explain why these plans are successful.

Going back to the Pareto filtered 31 920 plan set, we analyze groups of plans with the same ranking index and analyze the MOEs for the best, say 50 groups with highest mean ranking, which means the first approximately 2800 plans having mean ranks of at least approximately 0.43 in Fig. 7. We

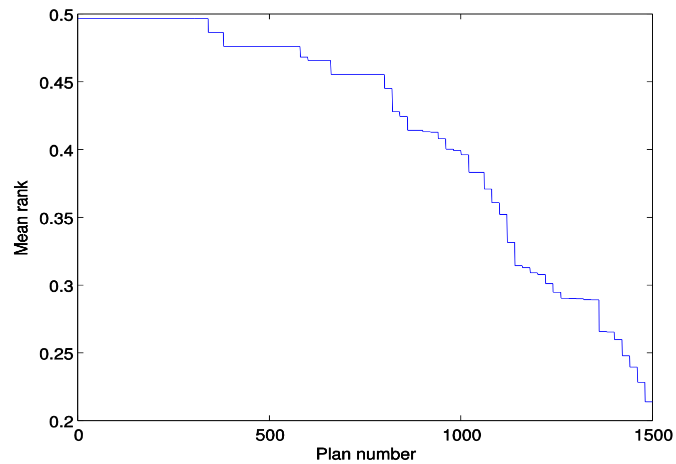

Fig. 9. The sorted mean normalized rank distribution of the Pareto-filtered plan set constrained to *BPI* = *RPI* = 7 platoons.

write *approximately* here since these values shift up and down slightly due to the Monte Carlo noise from simulation to simulation. The groups are numbered in the same order as the plan numbers in that figure (the best plans in the first group, the next best plans in group no. 2, etc.). The result for the different MOEs can be seen in Fig. 10.

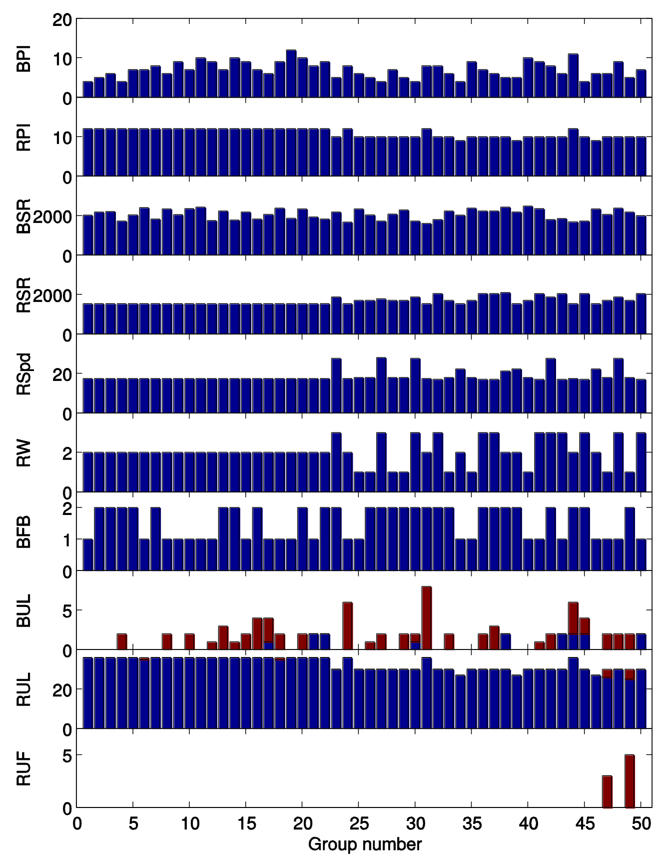

Fig. 10. The distribution of different MOEs for all Pareto-filtered plans depending on rank group, sorted downwards according to the mean rank. Blue bars are the *min* values of outputs from 20 different seeds to simulations with identical inputs for respective MOE, red bars marks difference between the *min* and *max* values.

The first seven are input parameters, followed by three output measures. Looking at these three measures we notice that BUL is the most important parameter to pay respect to, expected to be as low as possible. Thereafter, comes RUL, expected to be as high as possible. Blue losses BUL increases after the first twelve groups and red losses RUL start to decrease after the first 22 groups. The measure for red success *RUF* that could be argued to be the most important measure does not come into play for the first 46 groups.

Blue bars goes up to *min* values for varying seeds, and red bars mark the span up to the *max* levels. We see that bars are red-only for the best plans, so the simulation seed has here had much influence on BUL. We also note that the invariance for some parameters up to group no. 23 (approximately the best 1000 plans). We also notice that here $RUL \approx 3 \cdot RPI$ which means that almost all red vehicles are destroyed.

In this analysis, we have not fixed the sizes of the blue and red forces which might result in too many free input dimensions to analyze causality and find clear cause-consequence relations. We can use the above described grouping idea, but again constraining the number of platoons to *BPI* = *RPI* = 7 to observe the result for a more balanced situation (let be that blue platoons contain four vehicles and red platoons contain three). This let us focus on circa 1500 plans from the Pareto optimal frontier, see Fig. 9. We choose to focus on the 20 groups with highest mean rank, which contains approximately 1150 of the best plans with mean rank of at least approx. 0.335, see Fig. 11.

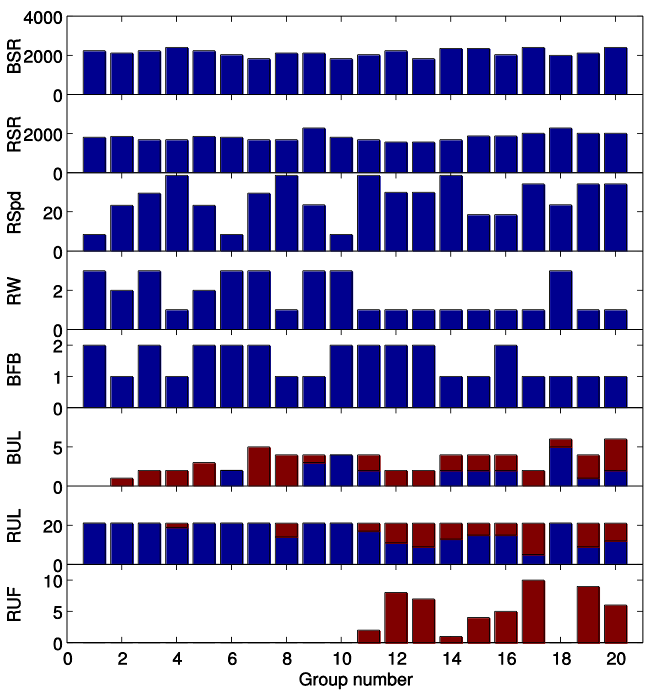

Fig. 11. The distribution of different MOEs for all Pareto-filtered plans constrained by *BPI* = *RPI* = 7, sorted in descending order by the mean rank. Blue bars are the *min* values of outputs from 20 different seeds to simulations with identical inputs for respective MOE, red bars marks the difference between the *min* and *max* values.

Since we have fixed *BPI* = *RPI* = 7, we have in the figure omitted the bar plots for those MOEs compared to Fig. 10. We now have much fewer plans to choose among (also reducing simulation fidelity), but we can see that the first group of plans, containing approximately 350 very similar plans, gives the best expected result; zero *BUL*, maximum *RUL* and zero *RUF*. From observing the five presented input parameters, the reason for success seems to lie in a combination of parameter values. For the first 10 groups *RW* is usually 3, *BSR* > *RSR*. *RSpd* varies, but is low for the best group of plans. From the red-only bars for *RUF* we also see that simulation seed has high importance for this MOE; for some seeds, no red vehicles succeed to break through, for others up to ten vehicles do.

VIII. Conclusions

We have derived a multi-process multiple-criteria decision making methodology for assessing military plans within defense planning. The plans are evaluated by several different measures of effectiveness by a simulation system. The decision support methodology uses Pareto analysis followed by preference analysis of measures of effectiveness and Monte Carlo weighting of the measures within the given preference order. By using this methodology, it is possible to get an estimate of ranking of each plan and further analyze the best plans to learn which parameters and combination of parameter ranges that leads to success.

Acknowledgment

We wish to express our appreciation to Prof. Rassul Ayani of the Royal Institute of Technology, Stockholm, for fruitful discussions.

References

[1] A. K. Sen, "Markets and freedom: Achievements and limitations of the market mechanism in promoting individual freedoms," Oxford Economic Papers, vol. 45, no. 4, pp. 519–541, October 1993.

[2] M. Ehrgott and M. M. Wiecek, "Multiobjective programming," in Multiple Criteria Decision Analysis: State of the Art Surveys, J. Figueira, S. Greco, and M. Ehrgott, Eds. New York: Springer, 2005, pp. 667–722.

[3] L. V. Utkin, "A new ranking procedure by incomplete pairwise comparisons using preference subsets," Intelligent Data Analysis, vol. 13, no. 2, pp. 229–241, April 2009.

[4] J. Schubert, "On ρ in a decision-theoretic apparatus of Dempster-Shafer theory," International Journal of Approximate Reasoning, vol. 13, no. 3, pp. 185–200, October 1995.

[5] S. Huang, X. Su, Y. Hu, S. Mahadevan, and Y. Deng, "A new decision-making method by incomplete preferences based on evidence distance," Knowledge-Based Systems, vol. 56, pp. 264–272, January 2014.

[6] Z. Yue, "An extended TOPSIS for determining weights of decision makers with interval numbers," Knowledge-Based Systems, vol. 24, no. 1, pp. 146–153, February 2011.

[7] A. P. Dempster, "A generalization of Bayesian inference," Journal of the Royal Statistical Society Series B, vol. 30, no. 2, pp. 205–247, 1968.

[8] A. P. Dempster, "The Dempster-Shafer calculus for statisticians," International Journal of Approximate Reasoning, vol. 48, no. 2, pp. 365–377, June 2008.

[9] G. Shafer, A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press, 1976.