Estimating Relative Combat Effectiveness Using Simulations

Per-Idar Evensen, Marius Halsør, Dan Helge Bentsen
Norwegian Defence Research Establishment (FFI)
Kjeller, Norway
per-idar.evensen@ffi.no, marius.halsor@ffi.no, dan-helge.bentsen@ffi.no

ABSTRACT

The Norwegian Defence Research Establishment (FFI) is often interested in assessing and comparing the relative combat effectiveness of different combat systems. For example, which one of two, or more, alternative combat systems that can best solve a specified selection of tasks or missions. In this context, a combat system can for example be a weapon system, a group of fighting entities, a force structure element, or a force structure.

There is no precise and unambiguous definition of combat effectiveness. However, combat effectiveness can in general be said to be a measure of a combat system's ability to solve a given task or mission, or a measure of how well a combat system solves a given task or mission. Combat effectiveness is affected by many different factors, and measuring and analyzing combat effectiveness is a complex and challenging task. Data from real-world warfare are often scarce, and for obvious reasons, it is of course not possible to experiment with warfare in the real world. Modeling and simulation (M&S) is therefore essential to experiment with, and assess and compare the performance and combat effectiveness of, different weapon systems and force structure elements. Combat models, however, are simplifications and will never represent all aspects of reality. Simulating modern warfare with sufficient realism is very challenging, especially when it comes to human factors.

In this paper we discuss what combat effectiveness means from a conceptual, system-theoretic perspective, look at some of the definitions of combat effectiveness, and discuss the factors affecting combat effectiveness. Moreover, we briefly describe and discuss some of the approaches for quantifying and measuring combat effectiveness that have been suggested in the literature. Finally, we present our general approach for using simulations and simulation-supported wargames to assess and compare the relative combat effectiveness of different combat systems.

ABOUT THE AUTHORS

Per-Idar Evensen is a Senior Scientist at the Norwegian Defence Research Establishment (FFI), where he has worked since 2006. His research focuses on M&S of land force operations at different levels for experimentation and analysis purposes, and he has a special interest in simulation systems based on game technology. He has a Master of Science degree in Computer Science from the University of Oslo in 2004.

Marius Halsør is a Principal Scientist at the Norwegian Defence Research Establishment (FFI), where he has worked since 1998. He has developed and utilized several simulation tools for experimentation and analysis for land force operations at tactical levels and below. He has a Master of Science degree from NTNU from 1997.

Dan Helge Bentsen is a Chief Research Engineer at the Norwegian Defence Research Establishment (FFI where he has worked since 2009. His work has mainly focused on the use of M&S tools to evaluate the combat effectiveness of different land force structures and components. He has served as a tank commander in the Norwegian Army and brings his experience from deployments to Afghanistan. He has a bachelor's degree in Media Technology and Design from the University of Agder.

Estimating Relative Combat Effectiveness Using Simulations

Per-Idar Evensen, Marius Halsør, Dan Helge Bentsen
Norwegian Defence Research Establishment (FFI)
Kjeller, Norway
per-idar.evensen@ffi.no, marius.halsor@ffi.no, dan-helge.bentsen@ffi.no

INTRODUCTION

The Norwegian Defence Research Establishment (FFI) is often interested in assessing and comparing the *relative combat effectiveness* of different combat systems. For example, which one of two, or more, alternative combat systems that can best solve a specified selection of tasks or missions. In this context, a *combat system* can for example be a weapon system, a group of fighting entities, a force structure element, or a force structure.

There is no precise and unambiguous definition of combat effectiveness. However, combat effectiveness can in general be said to be *a measure of a combat system's ability to solve a given task or mission*, or *a measure of how well a combat system solves a given task or mission*. Combat effectiveness is affected by many different factors, and measuring and analyzing combat effectiveness is a complex and challenging task. Data from real-world warfare are often scarce, and for obvious reasons, it is of course not possible to experiment with warfare in the real world. Modeling and simulation (M&S) is therefore essential to experiment with, and assess and compare the performance and combat effectiveness of, different weapon systems and force structure elements. Combat models, however, are simplifications and will never represent all aspects of reality. Simulating modern warfare with sufficient realism is very challenging, especially when it comes to human factors.

In this paper, we first describe the background for this work. After this, we discuss what combat effectiveness means from a conceptual, system-theoretic perspective, and look at some of the definitions of combat effectiveness. Furthermore, we discuss the factors affecting combat effectiveness, and briefly describe and discuss some of the approaches for quantifying and measuring combat effectiveness that have been suggested in the literature. Finally, we present and discuss our general approach for using simulations and simulation-supported wargames to assess and compare the relative combat effectiveness of different combat systems and outline some suggestions for further work.

BACKGROUND

An ever-recurring research question for FFI is: how to increase the combat effectiveness for different parts of the Norwegian Armed Forces? As part of this research, we assess and compare the performance and effectiveness of different combat systems, which vary with regard to composition of materiel and equipment (for example introduction of new technologies), tactical organization, or operational concept.

Data from modern real-world combat are often scarce and difficult to obtain, especially data encompassing the use of new technologies (e.g. new sensor systems, weapon systems, or protection systems) and new and novel operational concepts. Therefore, we use M&S as our primary tool for experimenting with, and analyzing the performance and effectiveness of, different combat systems.

Generally, we use *virtual* simulations in experiments where human system operators are essential, for example when experimenting with technologies directly affecting human performance or how humans operate at the technical level. An example is the experiments where we used virtual simulations to evaluate the operational benefit of augmented reality (AR) for improved situational awareness in combat vehicles by comparing the performance and effectiveness of combat vehicles with and without the AR system (Evensen & Halsør, 2013; Evensen & Halsør, 2014). For practical reasons, the sizes of our virtual simulation experiments have been limited to a few platoons (reduced company level).

To simulate larger operations, typically at the battalion and brigade level, we use *constructive* simulations with *semi-automated forces* (SAF). An example is the series of simulation-supported wargames conducted in FFI project "Future land forces", where the performance and effectiveness of five fundamentally different land force structures were tested in a set of chosen scenarios (Hoff et al., 2012; Hoff et al., 2013). The goal of the experiment series was to rank the force structures based on their relative performance and effectiveness.

We strive to collect as much data as possible from our simulation experiments, including both quantitative and qualitative data. Examples of quantitative data collected during a simulation session are entity positions, detections from sensors, weapons fired, and entities hit and killed. This type of data can for example be used to construct *kill matrices*. A kill matrix is essentially a matrix showing which entity types on one side killed which entity types on the other side. Examples of qualitative data collected are the observations made during the operational planning process before the simulated operation, observations made during the simulated operation, and notes from discussions with the operators/players at the after action review (AAR) session. These data then form the basis for the analysis, where we usually want to rank the tested combat systems based on their relative performance and effectiveness. This is not always an easy task. In some cases, when we assess and compare a set of combat systems and two, or more, of them solve the given tasks or missions, it is useful to have additional measures of performance (MOPs) and measures of effectiveness (MOEs) for ranking the combat systems. Furthermore, there is always a risk that the conclusions may be too much based on human judgement without having enough supporting data, or too much based on the kill matrices alone and thereby fail to take into account the complete picture of the battle. It would therefore be useful to have quantifiable MOPs and MOEs supporting the manual judgement and widen our understanding of the battle. For example, measures that can help explain why one combat system solved its task or mission and another combat system failed. Ideally, what we want is a clear definition of combat effectiveness and a way to measure this as quantitatively as possible. This paper is intended as a step towards this goal.

COMBAT EFFECTIVENESS

Combat effectiveness measures the quality of combat execution. In this section, we take a conceptual look at the evaluation of combat and measurement of combat effectiveness. Specifically, we discuss what combat effectiveness means from a system-theoretic perspective, look at some of the definitions of combat effectiveness, and discuss the factors affecting combat effectiveness. Finally, we describe and discuss some of the approaches for quantifying and measuring combat effectiveness that have been suggested in the literature.

Evaluation of Combat

As previously mentioned, our primary goal for analyzing modern warfare is to compare different combat systems or force structures, which may vary with regard to composition of materiel and equipment, tactical organization, or operational concept. To this end, it is useful to have a set of criteria and preferably a set of measurable parameters for ranking the quality of combat execution.

Economy, Efficiency and Effectiveness

If we look at a *system* in general, it is often useful to consider the three "E"s: *economy*, *efficiency* and *effectiveness*. A system usually has *inputs* and *outputs*. A useful system also has *outcomes*, which are the effects of the outputs on the external environment. The external environment is everything outside the boundary of the system.

For a system, *economy* relates to minimizing inputs, *efficiency* relates to the ratio of outputs to inputs, and *effectiveness* concerns the maximization of desired outcomes (Storr, 2009). This is illustrated in Figure 1. It is important to recognize that "[e]conomy and efficiency are not the same as effectiveness, and this is particularly true in combat" (Storr, 2009). Military units and combat systems need to be robust and have redundancy to survive, and therefore, to some extent, need to be inefficient and uneconomical in order to be effective.

A *system* can be defined as "[a] collection of components organized to accomplish a specific function or set of functions" (U.S. DoD MSCO, 2011). In this paper, we generally use the term *combat system* for any *system that has a function or set of functions in combat execution*. Examples of a combat system are a weapon system, a group of fighting entities, a force structure element, or a force structure.

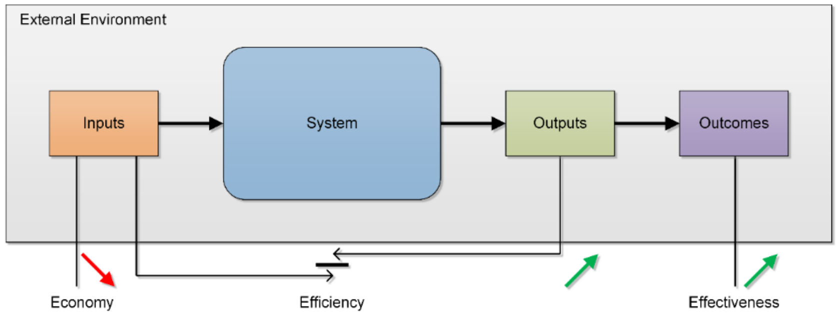

Figure 1. Economy, Efficiency and Effectiveness of a System (Storr, 2009).

Measures of Performance and Measures of Effectiveness

When assessing a combat system it is useful to develop specific *measures of performance* (MOPs) and *measures of effectiveness* (MOEs). MOPs generally have the following characteristics (National Research Council, 2013):

- They assess what a combat system achieves in terms of technical performance.
- They measure what a combat system is doing.
- They are typically quantitative.

Examples of MOPs are the number of enemy targets destroyed and the time it takes to advance to a phase line.

MOEs generally have the following characteristics (National Research Council, 2013):

- They assess the impact of the actions of a combat system on the effectiveness of achieving a task or mission.
- They assess changes in behavior, capability, or operational environment.
- They measure what is accomplished.
- They do not measure task performance.
- They are typically more subjective than MOPs and can be both qualitative and quantitative.

Examples of MOEs are mission accomplishment, casualty effectiveness (i.e. the ability to cause higher enemy casualties relative to own losses) and spatial effectiveness (i.e. the ability to advance, and take and hold territory). A system with high performance will naturally also have good prerequisites for achieving high effectiveness.

Definitions of Combat Effectiveness

There is no precise and unambiguous definition of *combat effectiveness*. NATO defines combat effectiveness as "[t]he ability of a unit or formation, or equipment to perform assigned missions or functions" (NSO, 2016). Technically, from a system-theoretic point of view, combat effectiveness is *a measure of the outcomes from a combat system*. That is, *a measure of the effects of the outputs from the combat system on the external environment* (see Figure 1). In more practical terms, we can say combat effectiveness is *a measure of a combat system's ability to solve a given task or mission*, or *a measure of how well a combat system solves a given task or mission*.

A key element here is the last part, *"…a given task or mission"*. This means we measure how effective a combat system is in this particular situation only. This cannot be transferred to other situations or scenarios. If we have two different combat systems, CS_A and CS_B, it may well be the case that CS_A performs better than CS_B in one scenario, while CS_B performs better than CS_A in another scenario. An overall or absolute combat effectiveness does not really make sense. This also means when comparing two different combat systems, it is vital to determine which scenarios these combat systems are expected to perform in, and measure and compare the combat effectiveness in the relevant scenarios.

It would be useful to be able to predict the combat effectiveness of a combat system. However, combat is not deterministic – at least not in practice. This means that it is not possible to find and know the states of the necessary set of initial conditions, and compose (and be able to execute) a simulation model with the necessary level of fidelity, which determines the exact outcome of a combat task or mission. The best we can hope for is to estimate the probabilities for a set of predefined categories of outcomes (for example success and failure). Since combat effectiveness is a measure directly tied to the outcome of a task or mission, it can in a predictive sense only be used to say something about the probability of a combat system to be able to solve a given task or mission.

Generally, if a combat system repeatedly has shown the ability to solve a given task or mission well, we can say it has high combat effectiveness for this task or mission. Then it is reasonable to expect that it will also have a high combat effectiveness, for the same type of task or mission, in the future. In (Hayward, 1968), Philip Hayward writes that "[i]t is reasonable to postulate that the greater the combat effectiveness of the unit under consideration, the greater the probability of achieving a given degree of success". Furthermore, he proposes the following definition of combat effectiveness: "combat effectiveness is the probability of success in combat operations" (Hayward, 1968).

A related term often used is *combat power*. NATO defines combat power as "[t]he total means of destructive and/or disruptive force which a military unit/formation can apply against the opponent at a given time" (NSO, 2016). Combat power is a measure of the quantity and quality of the elements of a combat system, whereas combat effectiveness is a measure of the quality of the actual combat execution. Combat effectiveness will of course depend on combat power. A combat system with high combat effectiveness can be expected to make better use of its combat power (Lawrence, 2017).

Factors Affecting Combat Effectiveness

In the literature, there are mainly two directions regarding which factors are important for combat effectiveness. The first direction looks at combat effectiveness mainly as a result of human factors. In (Brathwaite, 2018), for example, "skill to fight" and "will to fight" are considered the essential components of combat effectiveness. The other direction has a more holistic view of combat effectiveness and is concerned with all factors that can possibly affect the course of a battle, including both human and material resources. In this paper, we will follow the holistic view of combat effectiveness. Following Philip Hayward, we argue the factors affecting the combat effectiveness of a combat system essentially can be placed in the following main categories (Hayward, 1968):

- **Combat system (Blue force) capabilities:**
  This category includes all human and material resources of the combat system. These resources comprise both quantitative and non-quantitative factors. Human resources include number of personnel and their job specifications, experience, training level, and so on. Examples of material resources are vehicles, weapon systems, sensor systems, communication infrastructure, and ammunition. Important factors for the fighting entities are firepower, protection, and mobility. Included in this category are also the organization of units, doctrine, tactics, and human factors like leadership, "will to fight" (or morale), and cohesion. Essentially, this category represents the combat power of the combat system.

- **Enemy (Red force) capabilities:**
  This category includes all human and material resources of the enemy force. It includes the same factors as the previous category. Essentially, this category represents the combat power of the enemy force. An enemy force with much combat power is expected to be harder to fight, so the enemy capabilities will obviously affect the probability of a combat system to solve a given task or mission.

- **Environment:**
  This category includes terrain, vegetation, and human-built structures, in addition to climate and weather. It is not surprising the environment will affect the probability of a combat system to solve a given task or mission. Rain, snow, or fog are examples of weather conditions that may reduce a combat system's combat effectiveness.

- **Combat system (Blue force) task or mission:**
  This factor needs to be included, because it is not possible to determine if a given task or mission has been solved, unless it has been defined. Furthermore, a given combat system may have a high combat effectiveness for one task or mission, but a negligible combat effectiveness for a task or mission of another nature.

- **Enemy (Red force) task or mission:**
  The enemy mission will also affect the probability of the combat system to solve a given task or mission.

These categories, and how they interact, are illustrated in Figure 2.

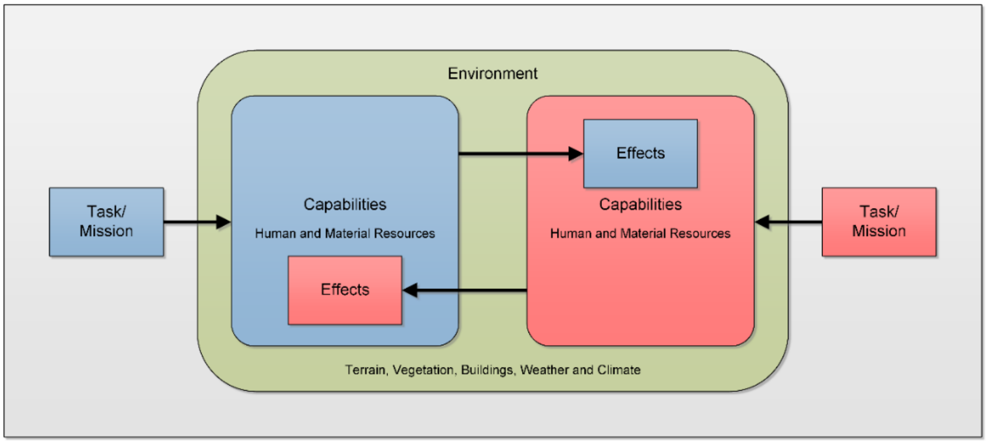

Figure 2. Factors Affecting Combat Effectiveness.

If we only look at the factors in the "combat system capabilities" category, and keep all the other factors constant, there are mainly three overall factors that determine the combat effectiveness:

- Force strength of the combat system relative to the enemy force
- Technological lead of the combat system relative to the enemy force
- Force employment of the combat system

Force employment appears to be the most decisive factor in modern warfare. In his book *Military Power: Explaining Victory and Defeat in Modern Battle* (Biddle, 2004), Stephen Biddle presents compelling arguments based on historical data that force employment has played a more important role for the outcome of warfare than either force strength or technological lead since the beginning of the twentieth century. Biddle states that effective force employment (which he calls modern system force employment) is "a tightly interrelated complex of cover, concealment, dispersion, suppression, small-unit independent maneuver, and combined arms at the tactical level, and depth, reserves, and differential concentration at the operational level of war" (Biddle, 2004).

How can Combat Effectiveness be Quantified and Measured?

There are three main approaches for measuring the combat effectiveness of a combat system:

1. Use data from actual combat situations.
2. Use a mathematical model of combat and solve it analytically.
3. Use data from simulations and wargames.

The only way of accurately measuring the combat effectiveness of a combat system is by using data from actual combat. However, as previously mentioned, data from actual combat are often scarce, and for the kind of studies we conduct, it is almost impossible to find useful data from actual combat. The second approach, deriving a useful mathematical model for combat and solving this analytically, is generally considered too complex. This leaves us with the last approach; estimating combat effectiveness by using data from simulations and wargames. We will discuss this approach further in the next section.

In the following subsections we will take a look at some of the approaches for quantifying and measuring combat effectiveness that have been suggested in the literature.

A Formal Equation for Combat Effectiveness

In (Hayward, 1968), Philip Hayward suggests the following expression for the combat effectiveness of a combat system:

$$\begin{aligned}\text{Combat effectiveness} &= \text{Probability of success in combat} \\ &= P(S) \\ &= f(x_1 \ldots x_n;\ y_1 \ldots y_n;\ e_1 \ldots e_m;\ m_1 \ldots m_r)\end{aligned} \quad (1)$$

where x_i = ith capability of the combat system, y_i = ith capability of the enemy force, e_i = ith environmental parameter, and m_i = ith mission parameter.

Hayward's expression for combat effectiveness, and his five main categories of factors affecting combat effectiveness, are useful as a conceptual framework. However, it is not possible to use it to quantify or measure combat effectiveness in practice. Hayward himself writes, "it is quite possible that even the most thoroughgoing analysis will leave a rather large number of independent variables, the influence of which on combat effectiveness remains to be assessed" (Hayward, 1968).

Loss Exchange Ratios

"The most commonly used metric of combat effectiveness is casualty effectiveness, which is the ability of one side to cause losses of another compared to their own losses" (Lawrence, 2017). The *loss exchange ratio* (LER) can generally be defined as the ratio of the number of fighting entities lost by the enemy divided by own losses. Since this ratio may vary throughout the battle, it is often useful to follow the development of the LER as a function of time into the battle:

$$LER(t) = \frac{R_0 - R(t)}{B_0 - B(t)} \quad (2)$$

where t = time into the battle, R_0 = initial number of Red entities, $R(t)$ = number of Red entities left at time t, B_0 = initial number of Blue entities, and $B(t)$ = number of Blue entities left at time t.

If we divide the losses on each side by the initial number of fighting entities, we get what is known as the *force exchange ratio* (FER) or *fractional exchange ratio*. The losses that are counted are usually personnel, but it can also be combat vehicles or other weapon platforms. Using weighted sums of the relative combat values of the entities from the different categories of fighting entities in use will probably give the best overall picture of the ratio of combat power that has been lost on each side. Historical data show that the percentage of losses for successful forces are almost always lower than the percentage of losses for their unsuccessful opponents, regardless of who is attacker and who is defender (Lawrence, 2017). The LER is therefore a well-founded measure for combat effectiveness. Nevertheless, LERs have limited usefulness regarding insights into the details and dynamics of the battle and the analysis of cause-and-effect relationships.

Battle Trace

In (Barr et.al, 1991), Barr, Weir & Hoffman suggest an approach for measuring combat effectiveness, which they call the *Battle Trace*. The Battle Trace indicates the degree of success of a force as a function of time into the battle. If the battle is divided into time intervals $\Delta_1, \Delta_2, \ldots, \Delta_n$, the Battle Trace for a given time interval Δ_i, is defined as (Barr et.al, 1991):

$$BT(\Delta_i) = \frac{\dfrac{R(t_i) - R(t_{i-1})}{R(t_i)}}{\dfrac{B(t_i) - B(t_{i-1})}{B(t_i)}} = \frac{\dfrac{\Delta R_i}{R_i}}{\dfrac{\Delta B_i}{B_i}} \quad (3)$$

where t = time into the battle, Δ_i = time interval from t_{i-1} to t_i, $R(t)$ = number of (effective) Red entities left at time t, and $B(t)$ = number of (effective) Blue entities left at time t. It is emphasized by Barr et al. that the number of forces used to calculate BT for a given time interval should include only entities that are able to affect opposing entities, for example by having line of sight and weapon range (Barr et.al, 1991). From equation (3) we can see that:

- If $BT < 1$ at a given time interval, the Red force is most successful.
- If $BT > 1$ at a given time interval, the Blue force is most successful.

The Battle Trace is a measure of which force that appears to be winning the battle at a given time. It is especially suitable for being used on data from detailed, entity-level combat simulations, where it can trace the course of the battle in real time. To use this approach in practice, however, clear criteria for which entities that should be counted must be defined. In addition, like with the LER, it would probably be better to use weighted sums of the relative combat values of the entities from the different categories of fighting entities in use.

Identifying and Counting Fire Engagement Opportunities

In (Lee & Lee, 2013) and (Lee & Lee, 2014), Lee & Lee suggest an approach for measuring combat effectiveness by identifying and counting fire engagement opportunities. They argue that combat effectiveness of a force under a direct fire engagement can be reasonably assessed by the number of attack opportunities it is able to create. An attack opportunity can either be created by a single entity, or by collaboration between two, or more, entities. Moreover, they model the combat environment as a *heterogeneous network* and use this representation to identify and count the occurrence of isolated and networked attack opportunities.

The concept of identifying and counting fire engagement opportunities is something that could be investigated further in general. It should be possible to implement an automatic system for counting fire engagement opportunities in combat simulations.

Combat Effectiveness Analysis Using Machine Learning

In (MacAllister et al., 2021), MacAllister et al. suggest an approach for using machine learning to determine the importance of different metrics for the outcome of a scenario. Specifically, the approach uses a Random Forest classifier to identify the most impactful effectiveness/performance indicators, followed by auxiliary analysis to determine in which direction the metrics drive the result.

This approach is interesting, and it is useful to be able to determine which factors that are most important for the combat effectiveness in a scenario. This approach, however, requires many simulation runs and therefore requires fully automated closed-loop simulations.

ESTIMATING COMBAT EFFECTIVENESS USING SIMULATIONS

As we concluded in the previous section, using M&S is the only feasible way in which we can study possible future combat scenarios. Especially when it comes to assessing the combat effectiveness of new and novel combat systems. For example, systems using new technologies, systems utilizing new ways of organizing units, or systems employing new concepts of operation. It is, however, important to keep in mind that combat models are simplifications and will never represent all aspects of reality. Military operations are highly complex and M&S of such operations, with sufficient realism, is very challenging, especially when it comes to the human factors.

In this section, we first discuss the two main approaches for combat simulations and their pros and cons. Then, we discuss how we use simulations to assess and compare the relative combat effectiveness of different combat systems and how to determine task or mission success. After this, we outline an example of a typical simulation experiment for assessing and comparing the relative combat effectiveness of different combat systems. Finally, we discuss the most important limitations of using combat simulations and simulation-supported wargames.

Combat Simulation Approaches

There are mainly two approaches for conducting combat simulations: using fully automated closed-loop simulations (without any human interaction) or using human-in-the-loop (HITL) simulations (with varying degrees of human interaction).

Closed-loop combat simulations can usually be run much faster than real-time. They can therefore be repeated hundreds or thousands of times to get a statistical distribution of the outcomes. Still, there is obviously a danger in removing the human elements entirely from the simulations since the human aspects are so important in combat, and human behavior is very difficult to model. Current computer-generated forces (CGF) can execute battle drills and lower-level tactics with a high degree of realism. However, it is currently not possible to create realistic models of the planning and decision-making (based on experience, intuition, and emotions) that are done by officers at the higher levels of the chain of command. In closed-loop combat simulations, the outcomes are only as good as the simulation model and associated input data themselves. Validation thus becomes very important, but combat simulations are in general very difficult to validate since it is most often not possible to compare the results to real-world situations.

Computer based HITL combat simulations are mainly associated with virtual simulations in the live, virtual, and constructive (LVC) taxonomy, but constructive simulations may also require a certain degree of human interaction, for example to control semi-automated forces (SAF). Including real humans in combat simulations is obviously the approach that gives the most realistic representation of the human aspects of combat. Using subject-matter experts (SMEs), like real system operators or real military commanders, in combat simulations will also function as an additional built-in face validation of the simulations. It is usually also easier to follow the course of the battle and identify cause-and-effect relationships in a HITL simulation, and thereby better understand what led to the outcome. Including military personnel, and especially military leaders, in the simulations, will also often lead to more confidence in the results among stakeholders (Hoff et al., 2012; Hoff et al., 2013). HITL simulations, and especially virtual simulations, are also more flexible when it comes to trying out new tactics and new ways of operating. HITL simulations are, however, time and resource consuming, and for virtual simulations it is often impractical, or not even technically possible, to use human operators for every entity in a simulation with hundreds of entities. In general, it is also problematic to execute HITL simulations faster than real-time, and this of course greatly limits the number of times a HITL simulation can be repeated. HITL simulations will also include the element of competition (since human participants generally don't like to lose). In a combat simulation with Blue and Red forces, both teams will do everything they can to win the battle, and the losing team will try to modify and improve their tactics for the next run.

In our simulation experiments, which are often focused on trying out new technologies and new ways of operating, we mainly use HITL simulations. As mentioned earlier, we use virtual simulations in experiments where human system operators are essential, for example when experimenting with technologies that directly affect human performance or how humans operate at the technical level. To simulate larger operations, typically at the battalion and brigade level, we use constructive simulations with SAF. These simulations are conducted as simulation-supported, two-sided (Blue and Red) wargames, where human players (typically officers) on both sides give orders to the SAF. Our methodology and best practices for conducting successful simulation-supported wargames for assessing force structures can be found in (Evensen et al., 2019) and (Evensen et al., 2021).

Relative Combat Effectiveness

In the discussions earlier in this paper, we argued that for a measure of combat effectiveness to be useful in general, it should be able to give an estimate of a combat system's probability of success in a given mission type. An absolute value for the combat effectiveness of a combat system, or the probability of a combat system to be able solve a given task or mission, would in practice be very hard to estimate with any particular confidence using simulations. Moreover, this estimate would in any case only be valid for a specific mission against a specific enemy in a specific environment.

Which one of two, or more, combat systems that has the highest *relative combat effectiveness* is, however, something that usually can be determined with more confidence using simulations than estimating an absolute value for combat effectiveness for each combat system. What we are mainly interested in, is to assess and compare the relative combat effectiveness of different combat systems executing the same task or mission against the same enemy. For example, to investigate which one of two, or more, alternative combat systems (e.g. weapon systems, groups of fighting entities, force structure elements, or force structures) that best solves a selection of tasks or missions. In our simulation experiments, we therefore seek only to vary the factors in the "combat system capabilities" category and keep all the other factors as constant as possible.

Determining Task or Mission Success

Battles are rarely fought until all units on one of the sides are completely destroyed. Usually, one side will reach some point where it no longer is able to reach its goal or function as a cohesive force. In a simulation-supported wargame the simulated operation will usually reach a point where there is a consensus among the umpires, and the Blue and Red players, as to whether the Blue combat system has successfully solved its task or mission or not. Additionally, quantitative measures like *breakpoints* can be used to determine success or failure. One approach for modeling breakpoints is to monitor the ratio between the remaining Blue and Red force and their initial force. When a certain percentage of one of the side's force has been lost, this force will no longer be able to continue the battle. This approach is known as the *absolute breakpoint rule* (Tolk, 2012). Another approach is to monitor the ratio between the Blue and the Red force in a battle. If this ratio reaches a level where one of the side's force is heavily outnumbered, this force may choose to retreat or surrender. This approach is known as the *proportional breakpoint rule* (Tolk, 2012). Figure 3 shows the development of the ratio of remaining Blue and Red forces in a battle until Blue wins as a result of the absolute breakpoint rule (to the left) and the proportional breakpoint rule (to the right).

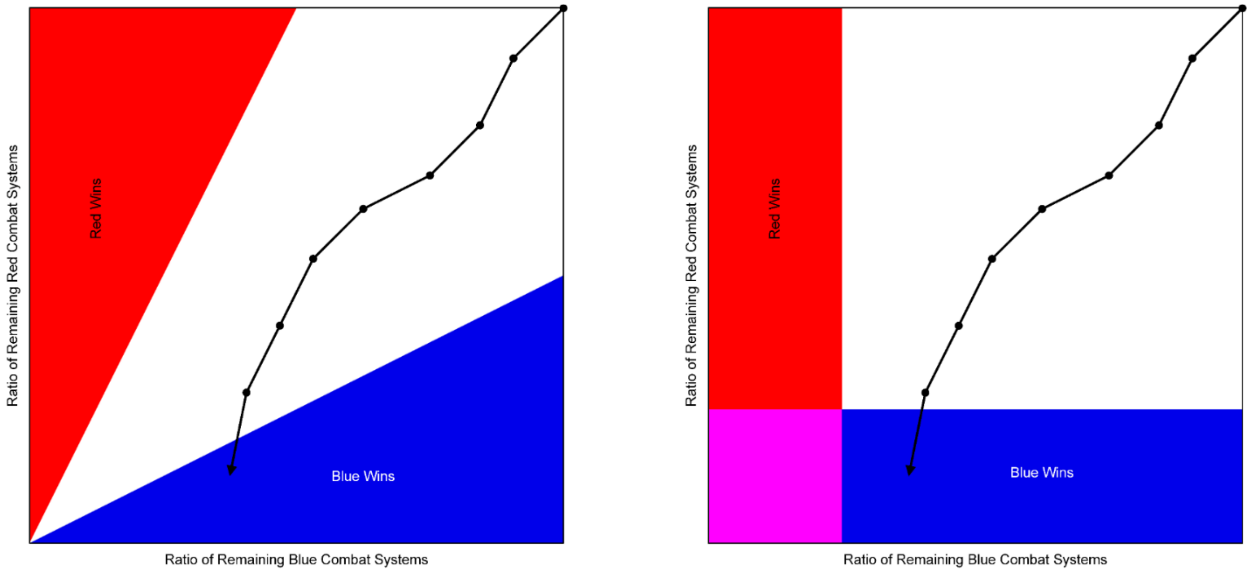

Figure 3. Absolute (to the left) and Proportional (to the right) Breakpoint Rule.

The loss exchange ratio (LER) and the Battle Trace approach will also give useful indications of which side is most successful. It is also important to note that more specific tasks or missions, for example in smaller virtual simulation experiments, can have other more specific qualitative or quantitative criteria for success. Criteria for task or mission success and specific MOPs and MOEs must be developed based on the combat system being evaluated and the given task or mission. An approach for a simulation experiment for comparing relative combat effectiveness is outlined in the following subsection.

Approach for Simulation Experiments

A typical research task for us can be to assess and compare the relative combat effectiveness of three different combat systems in three different scenarios. The three combat systems can for example consist of a combat system currently in use, CS_0, and two suggested alternative improvements (e.g. regarding materiel or concept of operation) of this combat system, CS_1 and CS_2. To test the combat systems we can for example choose three scenarios (or tactical vignettes). The scenarios should span the different tasks/missions the combat systems can be employed in. The scenarios we use are often derived from overarching national defense scenarios.

Depending on the level of human operator involvement necessary, and the size of the operation that needs to be simulated, the simulation experiments can be either virtual or constructive with SAF (i.e. simulation-supported wargames). It is also possible with simulation experiments that include both virtual and constructive entities. However, when virtual and constructive simulation models are combined, difference in resolution between them can lead to fair-fight issues (Evensen & Bentsen, 2016). Conducting virtual simulations to collect system performance data, and then using these data to calibrate subsequent constructive simulations, can also be an applicable approach.

Our primary indicator for combat effectiveness is a combat system's ability to solve the given task or mission based on given criteria. A typical result from a simulation experiment series with three combat systems tested in three scenarios can be in the form of a 3 x 3 matrix, where the cells are colored in green for success or red for failure. An example of such a result matrix is shown in Table 1.

In addition to being able to solve a given task or mission, it is, as mentioned earlier, often also useful to define other more specific MOPs and MOEs to evaluate the combat systems. In the example shown in Table 1, both CS_1 and CS_2 solves the given tasks or missions, but there could be a difference in how well or how effectively the tasks or missions are solved. Examples of additional criteria for success can be to complete the task or mission without losing more than 40 percent of the force or complete the task or mission within 24 hours. An example of a result matrix with an additional criterion is shown in Table 2. Here the cells are colored in yellow for task or mission success, but failure to fulfill the additional criterion. Furthermore, the cells are colored in green if the additional criterion is fulfilled. In this example, CS_1 fails to fulfill the additional criterion in two of the scenarios, while CS_2 fulfills the additional criterion in all three scenarios. This gives us an indication that CS_2 has the highest overall combat effectiveness of the three combat systems in the three chosen scenarios.

Table 1. Example of a result matrix that shows whether the combat systems solved the given task/mission.

	Scenario 1	Scenario 2	Scenario 3
CS_0	Failure	Failure	Failure
CS_1	Success	Success	Success
CS_2	Success	Success	Success

Table 2. Example of a result matrix that shows whether the combat systems solved the given task/mission and fulfilled the additional criterion.

	Scenario 1	Scenario 2	Scenario 3
CS_0	Failure	Failure	Failure
CS_1	Success	Success	Success
CS_2	Success	Success	Success

If both CS_1 and CS_2 fulfill (or fail to fulfill) the additional criterion, and no clear distinction in their combat effectiveness can be found, possible ways forward may be to repeat some, or all, of the scenarios or add new scenarios. The conclusion may also be that the simulation experiments indicate that their combat effectiveness is about equal. In any case, it is important to seek to understand why one combat system has a better relative combat effectiveness than the others do. For that reason, identifying major strengths and weaknesses of a combat system and its utilization is an important part of the analysis phase after an experiment series. The analysis phase may also result in suggested improvements that can further increase the combat effectiveness for the tested combat systems. Improved versions of the combat systems may then be tested in new simulation experiment series.

An example of a simulation experiment following the approach outlined above is the series of simulation-supported wargames conducted to estimate and rank the relative combat effectiveness of five fundamentally different land force structures (Hoff et al., 2012; Hoff et al., 2013). Another example is the virtual simulation experiments that compare the relative combat effectiveness for a platoon of combat vehicles with and without an AR system that visualized tactical graphics in the crew's sights and periscopes (Evensen & Halsør, 2013; Evensen & Halsør, 2014).

Limitations of Using Combat Simulations and Wargames

M&S is essential for most defense experimentation. Combat simulations and wargames can of course not be used to predict the exact outcome of a battle or a war, but they can produce plausible outcomes. As mentioned, it is challenging to validate combat simulations in general since it is most often not possible to compare the results to real-world situations. This becomes even more challenging for combat simulations that involve the use of future technologies and new and novel operational concepts. Even if the technical properties of the individual combat elements can be validated separately, combat is so complex, especially in its human aspects, that the best we can do to validate the simulation as a whole is to conduct face validation.

Today, it is possible to create synthetic environments that to a high degree replicate the physical properties of the real world. Modeling realistic human behavior and cognition, on the other hand, is recognized as the hardest and most complex challenge in combat simulation (Thorpe, 2010).

As we have argued, the combat effectiveness of a combat system depends on five main categories of factors (combat system capabilities, enemy capabilities, environment, combat system task or mission, and enemy task or mission). When using simulations to estimate relative combat effectiveness, it is important to consider if these categories of factors (especially the first three) are sufficiently represented in the simulations. Furthermore, it is important to consider what effects any weaknesses and limitations in the simulations can have on the results. For example, we have seen that entity-level simulation systems that do not represent micro-terrain features make cover and concealment difficult, and this systematically favors long-range, direct fire weapon systems (Evensen & Bentsen, 2016). It is also a known issue that current entity-level models tend to produce attrition levels that are higher than those observed historically because they lack good representations of the human aspects of combat and combat friction, resulting in that the simulated operations tend to run smoother than they would in the real world (Petty et al., 2012; Petty & Panagos, 2008).

We know that modeling realistic human behavior and cognition is very complex, and all simulation systems consequently have a limited representation of human factors. It is important to make an assessment of what effects this can have on the results.

FURTHER WORK

We need to continue our research to look for better ways to quantify and measure combat effectiveness. It could for example be interesting to further develop the Battle Trace approach by doing more research on developing criteria for which entities that should be counted. It could also be interesting to use the approach described by MacAllister et al. (MacAllister et al., 2021) to determine which factors that are most important for the combat effectiveness in specific scenarios, but then we would also need to start using fully automated simulations (as a supplement to our HITL simulations).

SUMMARY AND CONCLUSION

In this paper we have discussed combat effectiveness and how it can be quantified and measured. We have looked at some of the definitions of combat effectiveness, discussed the factors that can affect combat effectiveness, and described and discussed some of the approaches for quantifying and measuring combat effectiveness that have been suggested in the literature. Moreover, we have discussed how we use simulations to assess and compare the relative combat effectiveness of different combat systems.

Combat effectiveness can in general be said to be *a measure of a combat system's ability to solve a given task or mission*, or *a measure of how well a combat system solves a given task or mission*. The factors that can affect the combat effectiveness of a combat system can essentially be placed in the following main categories: combat system capabilities (human and material resources, organization, doctrine, tactics, etc.), enemy capabilities, environment (terrain, vegetation, human-built structures, climate and weather, etc.), combat system task or mission, and enemy task or mission.

There are three main approaches for measuring the combat effectiveness of a combat system: (1) use data from actual combat situations, (2) use a mathematical model of combat and solve it analytically, and (3) use data from simulations and wargames. However, data from actual combat are often scarce, and deriving a useful mathematical model for combat and solving this analytically, is generally considered too complex. The only feasible option for us, especially when it comes to assessing the combat effectiveness of new and novel combat systems, is therefore to use data from simulations and wargames.

An absolute value for combat effectiveness of a combat system would in practice be very hard to estimate with any particular confidence using simulations. Which one of two (or more) alternative combat systems that has the highest relative combat effectiveness, however, is something that usually can be determined with more confidence using simulation experiments.

ACKNOWLEDGEMENTS

The authors would like to thank our I/ITSEC Birddog Randy K. Jackson for his helpful comments and suggestions during the process of developing this paper.

REFERENCES

Barr, D., Weir, M. & Hoffman, J. (1991). *Evaluation of Combat*, Naval Postgraduate School, Technical Report NPS-MA-92-001.

Biddle, S. (2004), *Military Power: Explaining Victory and Defeat in Modern Battle*, Princeton University Press.

Brathwaite, K.J.H. (2018). Effective in battle: conceptualizing soldiers' combat effectiveness, *Defence Studies*, Volume 18, Issue 1.

Evensen, P.-I. & Bentsen, D.H. (2016). *Simulation of land force operations – a survey of methods and tools*, Norwegian Defence Research Establishment (FFI), FFI report 2015/01579.

Evensen, P.-I. & Halsør, M. (2013). Experimenting with Simulated Augmented Reality in Virtual Environments, Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2013, Paper No. 13028.

Evensen, P.-I. & Halsør, M. (2014). Using Virtual Environments to Evaluate the Operational Benefit of Augmented Reality, *Proceedings of the NATO Modelling and Simulation Group (NMSG) Annual Symposium 2014 (STO-MP-MSG-126)*, Paper No. 5.

Evensen, P.-I., Martinussen, S.E., Halsør, M. & Bentsen, D.H. (2019). Wargaming Evolved: Methodology and Best Practices for Simulation-Supported Wargaming, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2019*, Paper No. 19182.

Evensen, P.-I., Martinussen, S.E., Halsør, M. & Bentsen, D.H. (2021). *Simulation-supported wargaming for assessing force structures – methodology and best practices*, Norwegian Defence Research Establishment (FFI), FFI report 20/02595.

Hayward, P. (1968). The Measurement of Combat Effectiveness, *Operations Research*, Volume 16, No. 2, pages 314–323.

Hoff, E.Ø., Evensen, P.-I., Holhjem, H.R., Øyan, I.B. & Nygård, H.K. (2012). Simulation in Support of Army Structure Analysis, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2012*, Paper No. 12088.

Hoff, E.Ø., Evensen, P.-I., Holhjem, H.R., Øyan, I.B. & Nygård, H.K. (2013). Interactive Simulation to Support the Transition of Forces, *Proceedings of the NATO Modelling and Simulation Group (NMSG) Annual Symposium 2013 (STO-MP-MSG-111)*, Paper No. 6.

Lawrence, C.A. (2017). War by Numbers: Understanding Conventional Combat, Potomac Books.

Lee, Y. & Lee, T. (2013). A Measure to Assess Combat Effectiveness Using Network Representation, *Proceedings of the 2013 Summer Computer Simulation Conference (SCSC)*.

Lee, Y. & Lee, T. (2014). Network-based Metric for Measuring Combat Effectiveness, *Defence Science Journal, Volume 64*, No. 2, pages 115–122.

MacAllister, A., Rupp, P., Garrison, J., Hellstern, G., Javorsek, D. & Chu, P. (2021). Using Machine Learning for Battle Management Analysis, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2021*, Paper No. 21150.

National Research Council (2013). *Making the Soldier Decisive on Future Battlefields*, The National Academies Press.

North Atlantic Treaty Organization (NATO) Standardization Office (NSO). (2016). NATOTerm – The Official NATO Terminology Database. Retrieved March 31, 2022, from https://nso.nato.int/natoterm.

Petty, M.D., Franceschini, R.W. & Panagos, J. (2012). Multi-Resolution Combat Modeling, in *Engineering Principles of Combat Modeling and Distributed Simulation*, Tolk, A. (edited by), John Wiley & Sons.

Petty, M.D. & Panagos, J. (2008). A Unit-level Combat Resolution Algorithm Based on Entity-level Data. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2008*, Paper No. 8359.

Storr, J. (2009). *The Human Face of War*, Continuum.

Thorpe, J. (2010). Trends in Modeling, Simulation, & Gaming: Personal Observations About the Past Thirty Years and Speculation About the Next Ten, *Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2010*, Paper No. IF1001.

Tolk, A. (edited by) (2012). *Engineering Principles of Combat Modeling and Distributed Simulation*, John Wiley & Sons.

United States Department of Defense (U.S. DoD) Modeling and Simulation Coordination Office (MSCO) (2011). *Department of Defense Modeling and Simulation (M&S) Glossary*.