II. PREVIOUS STUDIES ON COMBAT MODELING

A. PREVIOUS STUDIES WITH LANCHESTER EQUATIONS

Past empirical validation studies of Lanchester Equations include the work of Bracken [Ref.8] on the Ardennes campaign of World War II, Fricker [Ref.6], also on the Ardennes campaign, Clemens [Ref.9] on the Battle of Kursk of World War II, and Hartley and Helmbold [Ref.10] on the Inchon-Seoul campaign of the Korean War. These works are among the few quantitative studies that use daily force size data for real battles.

1. Bracken's study

Bracken formulates four different models [Ref.8] for the Ardennes campaign, which are variations of basic Lanchester equations, and estimates their parameters for the first ten days of the of the Ardennes campaign of World War II (December 15, 1944 through January 16, 1945).

Bracken's models are homogeneous. Tanks, armored personnel carriers, artillery, and manpower are aggregated with weights representing the relative effectiveness of the weapon systems. This type of aggregation yields a single measure of strength for each of the Allied and German forces. This method is used to measure combat power and to calculate losses. His models treat combat forces and the total forces (i.e., both support forces and the combat forces) in the campaign separately.

Equations II.A.1.(3), II.A.1.(4) show the Lanchester equations used by Bracken, which are modified to include the tactical parameter d for Bracken's Model 1 and Model 2. The parameter d is a multiplier of attrition due to being either in a defensive or offensive posture in the battle. If $d < 1$, then the defender has fewer casualties (i.e., there is a defender advantage). If $d > 1$ then the defender has more casualties (i.e., there is an attacker advantage). If d=1 then there is no attacker or defender advantage. Using the tactical parameter *d* requires knowing which side is the defender and which side is the attacker.

$$\dot{B} = (d \text{ or} 1/d)\, a \; R^p B^q \qquad (3)$$

$$\dot{R} = (1/d \text{ or } d)\, b \; B^p R^q \qquad (4)$$

In Model 1, forces are composed of tanks, APCs, artillery, and combat manpower; where combat manpower is made up of infantry, armor, and artillery personnel. Manpower casualties are killed and wounded. Forces are tanks, APCs, artillery, and combat manpower, which are weighted by 20, 5, 40, and 1, respectively. That is, *Blue Forces (combat power) = (20* x *number of tanks) + (5* x *number of APCs) + (40* x *number of artillery) + (1* x *number of combat manpower).* Bracken [Ref.8] states in his study that, "The weights given above are consistent with those of studies and models of the U.S. Army Concepts Analysis Agency. Virtually all theater-level dynamic combat simulation models incorporate similar weights, either as inputs or as decision parameters computed as the simulations progress."

In Model 2, forces include all personnel in the campaign, including all types of logistics and support personnel. Casualties are personnel who are killed, wounded, captured or missing in action, and who have disease and nonbattle injuries. It is noteworthy here to mention that in the Ardennes campaign, the Allies had a smaller portion of their forces in combat units and a larger portion of their forces in logistics and support units than the Germans.

In estimating the parameters of Model 1, Bracken found that individual German effectiveness, as measured by the attrition parameter *a*, is less than Allied effectiveness *b*;

these parameters are for combat forces only. This distinction is a natural result of the German combat forces having less support, and therefore not being as effective as Allied combat forces individually. In Model 2 where all personnel are included, individual effectiveness is determined to be similar for both the Allied forces and the Germans.

In Model 3, the components used are the same as in Model 1, but the parameter *d* is not estimated. Just like Model 3, Model 4 does not have a tactical parameter. Model 4, like Model 2, addresses total forces rather than combat forces. For a summary of Bracken's models, see Table 2.

	COMBAT MANPOWER	SUPPORT MANPOWER	PARAMETER *d*
MODEL1	X		X
MODEL2	X	X	X
MODEL3	X		
MODEL4	X	X	

Table 2. Bracken's models summarized. Model 1 and Model 3 use combat manpower only; Model 2 and Model 4 use total manpower. Combat manpower is made up of infantry, armor, and artillery personnel; support manpower is made up of all types of logistics and support personnel. Model 1 and Model 2 have defensive parameter *d*; Model 3 and Model 4 do not have *d*.

Bracken's main conclusions are:

- Lanchester linear model best fits the Ardennes campaign data in all four cases.
- When combat forces are considered, Allied individual effectiveness is greater than German individual effectiveness. When total forces are considered, individual effectiveness is the same for both sides.
- There is an attacker advantage.

The second result indicates that the two sides have essentially the same individual capabilities but are organized differently. The Allies preferred to have more manpower in the support forces, which in turn yielded greater individual capabilities in the combat forces. The overall superiority of the Allied forces in the campaign led to the Allied attrition being a smaller percentage of their forces. Table 3 shows Bracken's best fitting parameters for the Ardennes campaign.

Name of the model	*a*	*b*	*p*	*q*	*d*
Bracken Model 1	8.0E-9	1.0E-8	1.0	1.0	1.25
Bracken Model 2	8.0E-9	8.0E-9	0.8	1.2	1.25
Bracken Model 3	8.0E-9	1.0E-8	1.3	0.7	-
Bracken Model 4	8.0E-9	8.0E-9	1.2	0.8	-

Table 3. Bracken's parameters found in his study for Ardennes campaign data.

2. Fricker's study

Fricker's paper [Ref.6] revisits Bracken's modeling of the Ardennes campaign of World War II [Ref.8] and uses the Lanchester equations. This is different than Bracken's study in several ways. Fricker's study:

- Uses linear regression to fit the model parameters.
- Uses the total body of data from the entire campaign, while Bracken used only the first 10 days of the data from the Ardennes Campaign.
- Also includes air sortie data.

In contrast to Bracken, Fricker shows that the Lanchester linear and square laws do not fit the data. He concludes by showing that a new form of the Lanchester equations—with a physical interpretation—fits best. Fricker states that the attrition parameter used in the Lanchester logarithmic model represents the opponent's probability of killing a soldier, and that this probability of kill is constant for a certain range of the opponent's force sizes. It follows that one side's losses are more a function of own forces rather than a result of the opponent's forces, and Fricker gives the Gulf War as support for this theory. That is, Iraqi casualties were more a function of the number of Iraqi forces than of the number of Allied forces. Table 4 shows the best fitting parameters for the Ardennes campaign according to Fricker's study.

Name of the model	***a***	***b***	***p***	***q***	***d***
Combat manpower w/o sortie	4.7E-27	3.1E-26	0.0	5.0	0.8093
Total manpower w/o sortie	1.7E-16	8.0E-16	0.0	3.2	0.824
Combat manpower With sortie	2.7E-24	1.6E-23	0.0	4.6	0.7971
Total manpower with sortie	1.3E-15	5.6E-15	0.0	3.0	0.8197

Table 4. Fricker's parameters from his study of the Ardennes campaign data. The estimated *d* parameter indicates a defender advantage. The *d* parameter used in Fricker's study is the inverse of the *d* parameter defined in Bracken's study.

3. Clemens' study

Clemens' analysis [Ref.9] examines the validity of the Lanchester Models as they are applied to modern warfare. The models in his study are based upon basic Lanchester Equations. The analysis is an extension of Bracken's [Ref.8] and Fricker's [Ref.6] analyses of the Ardennes Campaign, and applies the Lanchester models to the Battle of Kursk data.

Clemens uses two estimation techniques, linear regression and Newton-Raphson iteration. The analysis also explores the presented model in matrix form, and compares the matrix solution to the scalar solution. In his study he concludes that:

- Neither the Lanchester linear nor the Lanchester square model fits the data.
- The Lanchester logarithmic model in both scalar and matrix form fits better than the Lanchester linear and square models.
- Lanchester Equations do not give the best fit for the data.
- The analysis can be extended by:
  - Taking into account the change in offensive/defensive roles.
  - Adding data from air sorties.
  - Applying the Lanchester Equations in a homogeneous weapon scenario.
  - Building a whole new model without regard to the Lanchester formulations.

Table 5 shows the best fitting parameters Clemens found for the Battle of Kursk data in his study.

Name of the model	***a***	***b***	***p***	***q***	***d***
Clemens Linear Regression	6.92E-49	6.94E-48	5.3157	3.6339	-
Clemens Newton-Raphson	3.73E-6	5.91E-6	0.0	1.6178	-

Table 5. Clemens' parameters found in his study for the Battle of Kursk Data.

4. Hartley and Helmbold's study

Hartley and Helmbold's study [Ref.10] focuses on validating the homogenous Lanchester square law by using historical combat data. Since validating a model means testing it in a real life context, Hartley and Helmbold test Lanchester's square law using the data from the Inchon-Seoul campaign of the Korean War.

Hartley and Helmbold use three analysis techniques to examine the data; linear regression, the Akaike Information Criterion (AIC), and Bozdogan's consistent AIC (CAIC). The results of the study are:

- The data do not fit a constant coefficient Lanchester square law.
- The data better fit a set of three separate battles (one distinct battle every six or seven days). However, the data fit a set of three constant casualty-model battles just as well.
- Lanchester square law is not a proven attrition algorithm for warfare, but neither can it be completely discounted.
- More real combat data are needed to validate any proposed attrition law such as the Lanchester square law.

5. A summary of previous findings

Fricker's and Bracken's studies are significant in that they reach different conclusions using the same data. When both studies are compared, Fricker's approach and methodology makes more sense because he did not constrain himself to certain ranges of parameters, as Bracken did.

Bracken's approach is strong in the sense that his approach optimizes the nonlinear regression equation in the defined area. Fricker finds the parameters that give the minimum sum of squared residuals (SSR), using the logarithmically transformed Lanchester equations. Using logarithmic transformation does not necessarily guarantee the best fit when the parameters found by this approach are directly applied to the Lanchester equations. However, minimizing the SSR value was Bracken's criteria and the parameters found via logarithmic transformation in Fricker always resulted in smaller sums of square errors for the untransformed Lanchester equations than those found by Bracken.

In general, the results of all four studies show no overwhelming evidence of Lanchester fit. Among the three Lanchester equations, the logarithmic law gives the best fit.

B. THE DATA AND STUDY METHODOLOGY

1. The data

Complete combat data on both sides fighting against each other is very sparse. Consequently, validation of Lanchester and other combat models has been very difficult, and the most accessible battle data contains only starting sizes and casualties, sometimes only for one side. Furthermore, the definition of casualties varies (e.g., killed, killed plus wounded, killed plus wounded plus missing, killed plus wounded plus missing plus disease/nonbattle injuries), making data analysis difficult. Obtaining order-of-battle data and equipment damage reports requires extensive historical research. Recently, more data has become available and improved database management and computing power has helped in such data gathering efforts.

A detailed database of the Battle of Kursk of World War II, the largest tank battle in history, was recently developed. The data were collected from military archives in

Germany and Russia by the Dupuy Institiute (TDI), and are reformatted into a computerized data base, designated as the Kursk Data Base (KDB). The KDB was recently documented in the KOSAVE (Kursk Operation Simulation and Validation Exercise) study. [Ref.12]. The data are two-sided, time phased (daily), and highly detailed. They cover 15 days of the Battle of Kursk.

2. Study methodology

This thesis fits Lanchester equations and other functional forms to the newly released Battle of Kursk data. The two main areas of interest are the quality of the fits and the insights provided by the equations. Different fits are compared and contrasted to the previous research results mentioned above.

The methodology used in this thesis research consists of the following steps and research questions:

- Arranging and setting up of the data at hand so that it is useful for regression and statistical purposes.
- Conducting a thorough analysis and interpretation of the data.
- Identifying components needed for the model.
- Applying Bracken's and Fricker's methodology to the Kursk data.
- Applying various forms of Lanchester Equations to the data. How well do Lanchester Equations fit the Battle of Kursk Data?
    - Does the Linear Law fit the Battle of Kursk data?
    - Does the Square Law fit the Battle of Kursk data?
    - Does the Logarithmic Law fit the Battle of Kursk data?

- Do possible combinations of these three laws fit the Battle of Kursk data?

- Applying other general curve fittings and functional forms to the data.
- Do any of the other possible general curve fits or functional forms fit the Battle of Kursk data?
  - Do any of the functional forms need the defender/attacker coefficient?
  - What effect does changing weapon weights have on fitting the models to the data?
- Using a least squares grid search to get a better understanding of the relationship between various Lanchester formulation and the empirical data.
- Comparing and contrasting different methodologies and the two battles.
- Analyzing the results and conclusions of all the models.

This thesis extends the previous studies of Bracken, Fricker, Clemens, and Hartley and Helmbold in the following ways:

- Methodologies of previous studies are applied to Battle of Kursk data.
- A different regression technique, i.e., robust LTS regression, is used.
- Air sortie data is included.
- The change in offensive/defensive roles is taken into account.
- The battle is considered in different phases and different change points are used for fitting the model.
- Different weights are used for aggregating the data.

- Lanchester Equations, Morse-Kimball equations and force ratio models are fit to Battle of Kursk data
- Parameters found for different battles are used to fit Battle of Kursk data and the resulting parameters are compared and contrasted. By this comparison, the issue of whether or not the parameters of one battle can be used for another battle is discussed.