RAND

Cutting Some Trees to See the Forest

On Aggregation and Disaggregation in Combat Models

Richard J. Hillestad, Mario L. Juncosa

National Defense Research Institute

The research described in this report was sponsored by the Advanced Research Projects Agency. The research was conducted in RAND's National Defense Research Institute, a federally funded research and development center supported by the Office of the Secretary of Defense and the Joint Staff, Contract No. MDA903-90-C-0004.

Library of Congress Cataloging in Publication Data

Hillestad, R. J. (Richard John), 1942–
Cutting some trees to see the forest : on aggregation and disaggregation in combat models / Richard J. Hillestad, Mario L. Juncosa.
p. cm.
"Prepared for the Advanced Research Projects Agency."
"MR-189-ARPA."
ISBN 0-8330-1350-5
1. War—Mathematical models. 2. War games. I. Juncosa, Mario L. (Mario Leon), 1921– . II. United States. Defense Advanced Research Projects Agency. III. Title.
U21.2.H54 1993
355'.001'51—dc20 93-7406
CIP

RAND is a nonprofit institution that seeks to improve public policy through research and analysis. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published 1993 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
To obtain information about RAND studies or to order documents, call Distribution Services, (310) 393-0411, extension 6686

RAND

Cutting Some Trees to See the Forest

On Aggregation and Disaggregation in Combat Models

Richard J. Hillestad, Mario L. Juncosa

National Defense Research Institute

Preface

This report was prepared for the Advanced Research Projects Agency. The work was performed in the Applied Science and Technology program of RAND's National Defense Research Institute (NDRI), a federally funded research and development center sponsored by the Office of the Secretary of Defense and the Joint Staff. It is one of a trilogy of papers working through simple examples to illustrate deeper issues that arise in variable-resolution modeling. This report describes theoretical results regarding aggregation and disaggregation in combat models. The other papers are Paul Davis, *An Introduction to Variable-Resolution Modeling and Cross-Resolution Model Connection*, R-4252-DARPA, 1993, and Richard J. Hillestad, John Owen, and Donald Blumenthal, *Experiments in Variable-Resolution Combat Modeling*, N-3631-ARPA, forthcoming. Initial versions of all three papers were presented at a conference on variable-resolution modeling organized by RAND and the University of Arizona and sponsored by ARPA and the Defense Modeling and Simulation Office in May 1992.

Contents

Preface iii

Figures vii

Tables vii

Summary ix

Acknowledgments xi

1. INTRODUCTION 1

2. COMMON APPROACHES TO AGGREGATION AND DISAGGREGATION 3

3. AGGREGATION TO SCALARS: SCORING 4
   - Air/Ground Tradeoffs Using Scores 5
   - Disaggregation of Scored Results 6

4. THEORETICAL RESULTS ON AGGREGATION AND DISAGGREGATION 8
   - Consistent Scalar Aggregation 10
   - Partial Aggregation and Disaggregation of Lanchester Systems 14

5. CONCLUSIONS 19

Appendix: MATHEMATICAL CONSIDERATIONS 21

Figures

1. Alternative Disaggregations of Scored Combat 7
2. Consistency in Dynamic Models 8
3. Absolute Consistency in Aggregation 9
4. Consistent Scalar Aggregation Alternatives 14
5. An Example of Partial Aggregation 16

Tables

1. Consistent Scalar Aggregation 13
2. Consistent Partial Aggregation and Disaggregation 17

Summary

Most models of air and land combat use schemes of aggregation and disaggregation in representing combat systems, in spatial configuration, and in depicting the progress of a battle. For example, the use of firepower "scores" is an extreme case of aggregation of weapons into a single measure. Combining like systems into weapons categories—partial aggregation—is a common approach to representing a large number of aircraft or ground weapon types. This report explores different approaches to aggregation and what is known theoretically about aggregation and disaggregation in Lanchester combat models that in two dimensions are commonly called square-law models. It defines requirements for consistency between aggregate and higher-dimensioned models of this type. Some important conclusions are that aggregation should take into account the specific capabilities of the opponent (raising concern about many "scored" approaches that attempt to evaluate force components in isolation), and that partial aggregation (grouping "like" systems) and disaggregation of previously aggregated results can be done consistently only when certain restrictions on the relative attrition capabilities of weapon systems hold. When this is the case, specific nonarbitrary weightings can be determined for the partial aggregations.

Acknowledgments

The mathematical work in this report was stimulated by a preliminary investigation by N. Z. Shapiro. The authors acknowledge with gratitude the meticulous prepublication review by M. D. Miller.

1. Introduction

A common problem among military strategists and analysts is that of estimating the "strength" of a military unit. Such an estimate is needed to judge the success of operations, to compare the military forces of allies and opponents, and to determine "how much is enough" in defense budgeting. Yet military forces are composed of many distinct types of weapons and capabilities. Furthermore, success in battle is a function of many factors, including training, tactics, morale, terrain, command and control, etc. Ultimately a strength assessment is made on the basis of a *score,* that is, an aggregation of the force values into a single measure. Clearly, military experience and training permit some commanders to make good assessments of strength. But such estimates are nearly always highly situation-dependent, and are possible only if the commander has the required experience base.

When analysts evaluate new and perhaps undeveloped capabilities against possible threat forces, often they are forced to create an aggregate estimate of the strength of these forces. Thus, a fundamental problem of military analysis is how to aggregate or "score" a military force. Usually the weights come from what is considered experience, judgment, perhaps engineering or proving-ground tests, etc.

Aggregation is required for other reasons as well. Historical data are often available in less detail than might be desired; overall losses rather than losses of specific weapons systems may be all that is known, and certainly the exact causes of loss are usually unknown or unrecorded. Comprehension and understanding are frequently better served by more aggregate descriptions (the "forest versus trees" argument). Finally, analysis may require more efficient computation than is available with detailed battle simulations.

In general, the ability to aggregate and disaggregate combat forces and their processes is necessary and desirable—yet little theory and science underlie most approaches taken in this regard.[1] Aggregation in linear dynamical systems has been a topic of interest, e.g., in economics as well as in combat simulations, for

---

[1]See P. K. Davis and D. Blumenthal, *The Base of Sand Problem: A White Paper on the State of Military Combat Modeling*, RAND, N-3148-OSD/DARPA, 1991.

some time.[2] But the typical focus has been on nearly decomposable systems or weakly coupled systems that may be amenable to approximated methods for solution wherein the system is partitioned into smaller subsystems for individual treatment followed by a reassembly for solution of the given system.[3] However, the current authors are unaware of a systematic treatment of the conditions necessary to effect a consistent aggregation in the sense defined here. Combat models exhibit numerous approaches, but most are ad hoc. This report examines some of these approaches, illustrates some of the problems, and then attempts to describe what is known and possible in a theoretical sense for basic square-law Lanchester systems in dimensions higher than two.

The report is organized to first illustrate some common approaches used in aggregation or scoring, particularly in military conflict simulation models. Then, it shows some of the problems with these approaches. Finally, using Lanchester theory as a basis, it describes the requirements for theoretically "consistent" aggregation, disaggregation, and partial aggregation. An appendix provides the theorems and other mathematical considerations underlying the results in the main body of the text.

---

[2] H. A. Simon and A. Ando, "Aggregation of Variables in Dynamic Systems," *Econometrica*, Vol. 29, pp. 111–138 (reprinted in A. Ando, F. M. Fisher, and H. A. Simon, *Essays on the Structure of Social Science Models*, MIT Press, Cambridge, Massachusetts, 1963, pp. 64–91).

[3] G. Kron, "Solving Highly Complex Elastic Structures in Easy Stages," *Journal of Applied Mechanics*, Vol. 22, 1955, pp. 235–244.

2. Common Approaches to Aggregation and Disaggregation

Consider a military force composed of multiple types of weapons with different capabilities—the "combined arms" army, for example. Each weapon type and supporting system is important to prevent weakness (armor without infantry, for example), or because of dependencies (artillery requires adequate fire control), or to provide synergism in battle (rolling artillery barrages to reduce the defender advantage as attacking armor moves into range). Yet these weaknesses, dependencies, and synergisms are influenced by terrain, opponent capability, and tactics to be employed. How can one aggregate or score such a situation-dependent process?

The simplest form of aggregation involves assigning a value to each weapon type, multiplying by the number of each type and adding these values to obtain a force score or *total aggregation*. This is the approach taken by the WEI/WUV method,[1] which, although once used widely, is not currently favored by the U.S. Army as an approach to modeling. Nevertheless it is still used in various models and academic debates.[2] Some uses of scoring have gone further. For example, scores have sometimes been *disaggregated* and the results used to estimate the losses of individual weapons systems or weapon categories, and the ratios of force scores have been used to predict movement of forces in combat.[3] Almost all evaluations of military capability in models or exercises use some form of aggregation because the number of different systems is too large to consider directly.[4] The number of different systems in a typical mechanized army division may be 25 or more. Thus, *partial aggregation* of similar systems is a common approach. The subject of this report is what can be said about the theoretical "correctness" of these various approaches to aggregation and disaggregation. We first take up the issue of total aggregation.

---

[1] *Weapon Effectiveness Indices/Weighted Unit Values III (WEI/WUV III)*, U.S. Army Concepts Analysis Agency, November 1979.

[2] For example, some models used in the public debate on conventional arms control in Europe and to estimate outcomes prior to the Desert Storm operations used scored forces. See J. Bracken, "Stability of Ground and Air Forces Without and With a Buffer," *Phalanx*, Vol. 24, No. 2, June 1991.

[3] B. W. Bennett, C. Jones, A. Bullock, and P. Davis, *Main Theater Warfare Modeling in the RAND Strategy Assessment System (3.0)*, RAND, N-2743-NA, 1988.

[4] Ibid.

3. Aggregation to Scalars: Scoring

Scores or strength estimates have been necessary from the beginning of organized combat. Initially, with largely homogeneous forces, such a score could be based on the quantity of personnel available and an estimate of individual strength. Even then, the number that could engage at any one time was important, and tactics and terrain could allow a smaller force to defeat a larger one. As forces became more heterogeneous it was necessary to give a relative evaluation of various components. The Soviets carried this to a scientific extreme with the "Correlation of Forces" methodology,[1] which attempted to predict the success of operations by evaluating force scores against time, position, and attrition objectives. A U.S. counterpart to this approach is the "Quantified Judgment Model" of Colonel T. N. Dupuy,[2] which evaluates the outcome of battles based on a force score that is situation dependent and draws data from historical battles. The WEI/WUV scoring approach estimates strength in terms of division equivalents, and a more recent approach called "Situation Adjusted Scores"[3] attempts to estimate force strength with more attention to the situations of combat. Various models attempt to use these force aggregations for purposes of predicting campaign outcomes in terms of attrition, force movement, and battlefield success.[4]

A number of problems arise in attempting to score complex military forces. First, the value of a force should depend strongly on the opponent and situation. A technically sophisticated armored force may not be of great military value in low-intensity urban or jungle conflict; advanced armored forces may be sitting ducks for an uncontested air force. The situation and opponent aspects of scoring may have been implicit during the Cold War, when the opponent was assumed to be the Soviets and the Warsaw Pact in Central Europe. But when the opposition could be any of a number of possible enemies with widely varying capability, one scenario's aggregations would seem to be inappropriate for many others. Clearly, Soviet tanks in the hands of Iraqi soldiers during the recent Gulf War did

[1] J. Hines, "Calculating War, Calculating Peace: Soviet Military Determinants of Sufficiency in Europe," in Reiner Huber (ed.), *Military Stability*, nomos Verlagsgesellschaft, Baden-Baden, 1990.

[2] T. N. Dupuy, *Understanding War*, Paragon House Publishers, New York, 1987.

[3] P. Allen, *Situational Force Scoring: Accounting for Combined Arms Effects in Aggregate Combat Models*, RAND, N-3423-NA, 1992.

[4] See, for example, the RSAS description referenced above, or Joshua M. Epstein, "The 3:1 Rule, the Adaptive Dynamic Model, and the Future of Security Studies," *International Security*, Spring 1989.

not have the lethality (score) that they might have had if operated by trained Soviet or East German soldiers.

Another problem is that determination of scores by subjective judgment has been unscientific to say the least. Experts, when called upon, must draw on experience. But the experience base is limited either to specific contrived exercises that cannot be expected to represent warfare realistically, or to historical experience with known weapons and situations. If experts are to evaluate proposed or experimental weapon systems or new situations and opponents, they must do it with guesswork.

As if the aggregation of a combined arms force were not difficult enough, it is even harder to use the aggregation to predict more than the likelihood of a single battle outcome. This requires a prediction of movement of forces in combat and, in order to predict the next battle, the composition of forces surviving. This means that the attrition results of earlier battles must be disaggregated. In mathematical parlance, this is a one-to-many mapping, and it simply cannot be done uniquely without additional information. Given a 10 percent attrition of an aggregated force, what components of the force survive? Should the losses be distributed evenly according to number of systems, or should they be distributed by relative vulnerability, relative lethality, or what? The following examples illustrate some of the problems with aggregation to scores and disaggregation.

Air/Ground Tradeoffs Using Scores

The first example is a simplification of a serious debate that took place during the development of NATO's conventional arms control position in the late 1980s.[5]

The basic question was whether tactical aircraft helped or hindered stability in the central region conventional-force balance. It is assumed that defensive predominance is good and attacker predominance bad from the standpoint of stability. Let A be the attacker ground force score and D be the defender ground force score. The ground force ratio is then $FR = A/D$. In one analysis, the air forces are added and considered to be killers of ground forces, and therefore they reduce the number of ground weapons and resulting scores. Assuming equal air forces for both attacker and defender, the forces removed by air attacks from each side is a. The resulting force ratio is then

$$FR = \frac{(A - \mathrm{a})}{(D - \mathrm{a})},$$

[5] Bracken, op. cit. See also the comments following Bracken's article in the same journal.

which favors the attacker under the doctrinal assumption that, as the attacker requires the greater force, we have $A > D$.[6] In this evaluation the air forces are considered to be *destabilizing*, as they permit the attacker to gain more of an advantage.

In a countervailing evaluation, the air forces are added and are considered to add firepower to each side. Thus, again assuming equal air forces, the firepower added is b to each side. The resulting force ratio is then

$$FR = \frac{(A + b)}{(D + b)},$$

which favors the defender under the same assumption that $A > D$, and in this evaluation tactical aircraft could be seen to be *stabilizing*. A similar paradox is easily developed for helicopters, long-range ground fires, etc. The important point is that the use of aggregation (of air power to ground forces) has somehow eliminated information that might be necessary in the evaluation of airpower and stability.

Disaggregation of Scored Results

There are two commonly used approaches to disaggregation in combat models. The simplest approach apportions the losses of individual weapon systems based on the initial proportion of the systems of each type. This means that a 10 percent loss in aggregated score causes a 10 percent loss in the quantity of each type of system. This type of disaggregation is illustrated in the left curve in Figure 1, which is based on a square-law Lanchester attrition calculation described in the next section. Note that this type of disaggregation keeps the proportion of weapons constant in time. On the right side of Figure 1 the disaggregation has been done in proportion to the relative weight of the particular weapon used to create the score. That is, let w_1 and w_2 be the weights ("scores") of two weapon systems and let $N_1(t)$ and $N_2(t)$ be the number of those systems at time t. The score of this two-weapon unit is

$$S(t) = w_1 N_1(t) + w_2 N_2(t).$$

Let the losses of this scored unit in an interval dt be denoted by dS, and let these losses be computed by the same Lanchester square law. In this type of disaggregation, the fractional losses of systems 1 and 2 in the interval dt are computed to be

[6] We assume that $A > D$, otherwise the potential attacker would probably not attack.

$$\frac{dN_1}{N_1(t)} = \frac{w_1}{w_1 + w_2} \frac{ds}{S(t)}$$

and

$$\frac{dN_2}{N_2(t)} = \frac{w_2}{w_1 + w_2} \frac{ds}{S(t)} .$$

After extracting these losses for the interval, the system is rescored and the Lanchester square law attrition is recomputed for the next time interval. This results in a disproportionate drawdown and, of course, a different overall solution. The weaker component of the forces takes a proportionately larger amount of the loss as time advances. Either case could occur. Without additional information about tactics, how fire was allocated, etc., it is not possible to say which, if either, is correct. This is an illustration of the one-to-many mapping problems that are prevalent in attempts to disaggregate results in combat models. When these results are used in scoring the forces in the next stage of battle, the predicted campaign outcomes can be dramatically different depending on the approach taken.

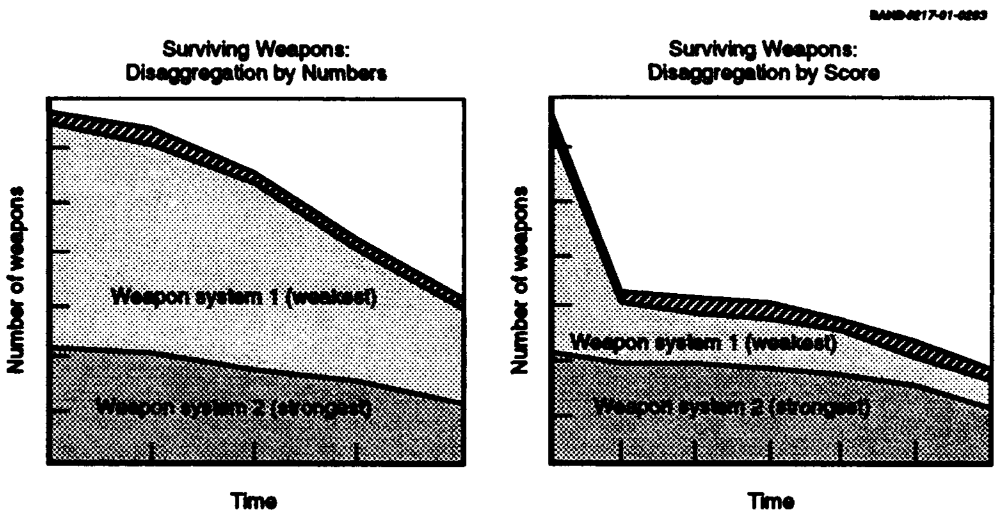

Figure 1—Alternative Disaggregations of Scored Combat

4. Theoretical Results on Aggregation and Disaggregation

Some important questions about aggregation and disaggregation are: How does an aggregation of a system relate to its more detailed representation? Given consistency requirements between two representations at different resolution, is there a "correct" way to aggregate? When is a consistent *partial* aggregation possible? When can aggregated combat results be disaggregated?

It is necessary to define the notion of consistency first. Figure 2 demonstrates the consistency between two models as defined in terms of the model output measures. At any time of interest the analyst should be able to compare the outputs of two models, and after an appropriate "mapping," the outputs should match within some small amount. The mapping is required because the analyst may be interested only in some aggregated outputs, and on the other hand, the outputs of one model may need to be converted into the similar measures produced by the other model. The differences between the outputs, because they represent a vector, must be measured in terms of a scalar norm of the differences. If this norm is small, then the models can be considered to have epsilon consistency with respect to the measures at the times of interest.

Figure 3 illustrates a slightly simpler representation of consistency. In this case we desire absolute consistency ($\varepsilon = 0$) between an aggregated model and a more

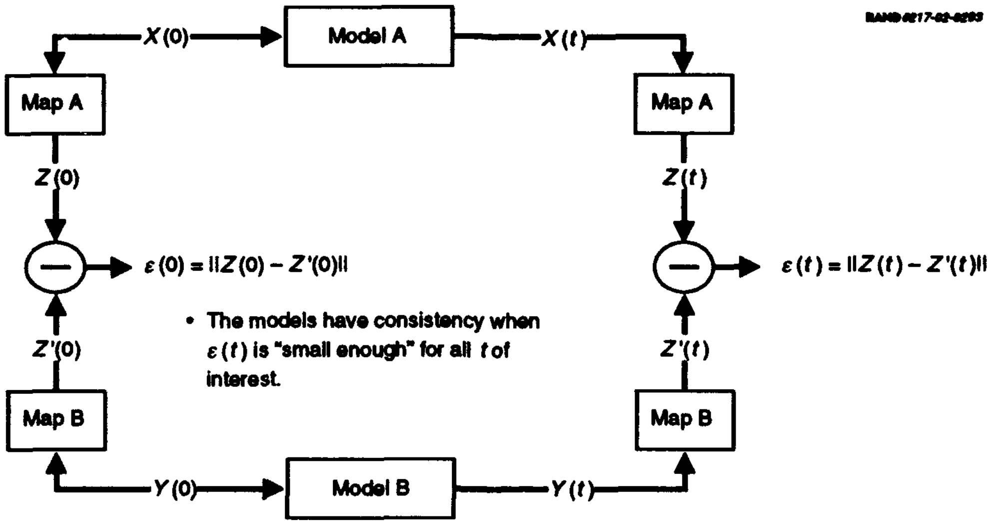

Figure 2—Consistency in Dynamic Models

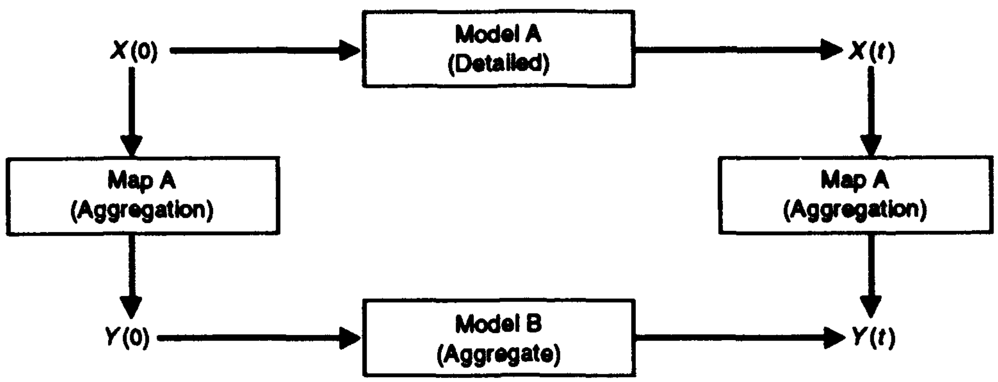

Figure 3—Absolute Consistency in Aggregation

detailed model. The mapping goes only one way—the outputs of the detailed model are aggregated to match the aggregate model. When this mapping can be done with the same mapping or aggregation functions at all times and the outputs match exactly, then we say the two representations have *absolute consistency* and are *commutative*. We now ask, under what conditions can we aggregate Lanchester square-law combat models partially and completely and maintain absolute consistency?

Because the constant-coefficient, heterogeneous Lanchester square-law[1] models lead to a system of linear differential equations, it is possible to derive and state strict theoretical results with respect to aggregation and disaggregation.[2] The square-law Lanchester system is described by the pair of vector differential equations

$$\frac{dX(t)}{dt} = -AY(t)$$

and

[1]See J. G. Taylor, *Force-on-Force Modeling*, Military Operations Research Society of America, Arlington, VA, 1981.

[2]It is clear that combat does not strictly follow a Lanchester square law, but understanding the requirements for aggregation and disaggregation of this "ideal" system is an important first step to understanding how it might be done in more complex models, say, possibly in ones represented by nonlinear differential equations that may be equivalent to higher-order linear ones, as one-dimensional and some higher-dimensional Riccati equations are. This matter requires further study.

$$\frac{dY(t)}{dt} = -BX(t) ,$$

where $X = [X_1, X_2, \ldots, X_m]$ and $Y = [Y_1, Y_2, \ldots, Y_n]$ are the vectors of side X and side Y weapons systems. $A = [A_{ij}]$ and $B = [B_{ij}]$ are the Lanchester coefficient matrices defining the rate at which Y systems destroy X systems and vice versa, respectively.

An *aggregation* of this system is a reduced-dimension system

$$\frac{dU(t)}{dt} = -CV(t)$$

and

$$\frac{dV(t)}{dt} = -DU(t) ,$$

where $U(t)$ and $V(t)$ are aggregations of X and Y such that $U = RX$ and $V = SY$. The vectors U and V are of length r and s which are less than m and n (otherwise, U and V would not be aggregations, according to our definition). R and S are aggregation mappings (matrices) which are nonnegative (it's not clear what negative weights of weapons systems would imply). Thus, if R is a matrix of dimension $1 \times n$, then the resulting U is a scalar. If R is $2 \times n$, then the vector of X systems is reduced to two aggregate components that comprise U. It is not necessary that X and Y or U and V have the same dimensions. U might be a scalar in the aggregated system and V a vector of systems.

Consistent Scalar Aggregation

Consider the case in which the aggregation matrices R and S are vectors and therefore map X and Y to scalars. This, in effect, weights the components for X and Y into scalar "scores" for the two sides. Results from linear algebra and systems theory dictate the requirements on R and S such that consistency between the resulting models is achieved. By consistency we mean that the result obtained from aggregating the solution of the unaggregated differential equation system is the same as the solution of the system of aggregated differential equations; i.e., in mathematical parlance, the operations of differential equation system solving and aggregation commute. We state the results here and provide the mathematical propositions and their proofs in the appendix.

First, the eigenvalues of the product matrices AB and BA are desired to characterize the time response of the systems. That is, the time response for X and Y is the solution of the system of differential equations as a function of time, subject, of course, to initial conditions $X(0) = X_0, Y(0) = Y_0$,

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} F(AB) & -G(AB)A \\ -G(BA)B & F(BA) \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix},$$

where

$$F(z) = \sum_{k=0}^{\infty} t^{2k} z^k \,/\, (2k)!$$

and

$$G(z) = \sum_{k=0}^{\infty} t^{2k+1} z^k \,/\, (2k+1)!\,,$$

both analytic functions over the finite complex plane and recognizable as $\cosh(t\sqrt{z})$ and $(\sinh(t\sqrt{z}))\,/\,\sqrt{z}$ respectively. Now, if we replace AB and BA by similarity transformations on their respective Jordan forms,

$$AB = W\Lambda W^{-1} \qquad \text{and} \qquad BA = Z\Gamma Z^{-1},$$

where the diagonal of Λ contains the eigenvalues of AB and the diagonal of Γ contains the eigenvalues of BA, we have a more computationally palatable form:

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} WF(\Lambda)W^{-1} & -WG(\Lambda)W^{-1}A \\ -ZG(\Gamma)Z^{-1}B & ZF(\Gamma)Z^{-1} \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix}.$$

Fact. ***The matrices* AB *and* BA *are nonnegative matrices with nonnull columns and nonnull rows and, consequently, have maximal positive eigenvalues and corresponding nonnegative left eigenvectors whose components are the weights that yield commutative or consistent scalar aggregations of* X *and* Y.**[3]

---

[3]The reason the product matrices, AB and BA, are important is seen by differentiating the equation for $dX(t)/dt$. This gives $d^2X(t)/dt^2 = -AdY(t)/dt = ABX(t)$, with a similar argument for BA. Thus, the solutions depend on AB and BA, rather than on A and B separately.

It is best to give an example of such an aggregation. Table 1 illustrates a consistent aggregation for a 2 × 2 Lanchester system reduced to a single component on each side. The original A and B matrices are shown at the upper left; possible C and D matrices (scalars in this case) are shown with the weights derived from the eigenvectors. There are two sets of weights because there are two possible eigenvalues and associated vectors.[4] The left part of the table shows the direct integration of the heterogeneous system in X and Y as a function of time; the columns labeled "weighted" are obtained by weighting the timewise values of X and Y. The columns labeled "integrated" $u(t)$ and $v(t)$ are obtained by using the R and S obtained with the left eigenvalues and integrating initial values $u(0)$ and $v(0)$. The fact that these match at each time t shows the consistency of the aggregation. The columns labeled "$ur(t)$" and "$vr(t)$" demonstrate that consistency is also achieved with weights based on the other (right) eigenvectors.[5]

Consider what has been lost by the aggregations when one chooses such a consistent set of weights. The two-dimensional Lanchester system in Table 1 has only positive components on the diagonals. These are actually the equations for two separate battles. X_1 shoots at Y_1 and Y_1 shoots at X_1. X_2 shoots at Y_2 and Y_2 shoots at X_2, but there are no shots between X_1 and Y_2, X_2 and Y_1, or vice versa. As a result of the preceding, we have that the consistent aggregations weight the components of X with $R = [0\ 1]$ or $R = [1\ 0]$ and, similarly, $S = [0\ 1]$ or $S = [1\ 0]$. Thus, either one battle or the other can be represented in a consistent aggregation, and there is no consistent aggregation for an "average" battle incorporating components of both.[6] The columns labeled "*uu*" and "*vv*" represent an attempt to approximate an average battle by averaging the weights. Comparison of the columns labeled "weighted" with those labeled "integrated" indicates that consistency has not been maintained in this attempt.

The point of this example is to show that something is *lost* in an aggregation, namely some aspect of the response of the detailed system. And as the illustration shows, the components left off may be important (in this case, the other battle). Figure 4 illustrates this example, showing the aggregated response

---

[4]Consistency applies only as long as no force components become negative. When this happens the original systems and aggregations must be changed, as negative weapons have *no* meaning. In the original systems, one simply deletes those components of X or of Y that go to zero and the corresponding rows and columns in A and B and proceeds with the reduced system of equations with, of course, new aggregation operators R and S.

[5]There is an optional scale factor that can also be applied to change the overall magnitude of the aggregated u and v components.

[6]This can be understood intuitively by considering the fact that the different components will generally attrite at different rates, leading to a continuously changing mix of systems.

Table 1

Consistent Scalar Aggregation

A Matrix

	1	2
1	0.50	0.00
2	0.00	0.20

B Matrix

	1	2
1	0.50	0.00
2	0.00	0.20

$dX/dt = -AY$
$dY/dt = -BX$

x weights (left)	
0.00	1.00
y weights (left)	
0.00	1.00

a (left)	0.2

b (left)	0.2

$u = w_1^* x_1 + w_2^* x_2$
$v = w_1^* y_1 + w_2^* y_2$

$du/dt = -a^* v$
$dv/dt = -b^* u$

x weights (right)	
1.00	0.00
y weights (right)	
1.00	0.00

a (right)	0.5

b (right)	0.5

					Aggregation with left eigenvalue				Aggregation with right eigenvalue				Aggregation with average weight			
					Weighted		Integrated		Weighted		Integrated		Weighted		Integrated	
t	$x_1(t)$	$x_2(t)$	$y_1(t)$	$y_2(t)$	$u(t)$	$v(t)$	$u(t)$	$v(t)$	$ur(t)$	$vr(t)$	$ur(t)$	$vr(t)$	$uu(t)$	$vv(t)$	$uu(t)$	$vv(t)$
0.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.3	1000.3	1000.3	1000.3	1000.0	1000.0	1000.0	1000.0
0.1	950.0	980.0	950.0	980.0	950.0	950.0	950.0	950.0	980.3	980.3	980.3	980.3	965.0	965.0	965.0	965.0
0.2	902.5	960.4	902.5	960.4	902.5	902.5	902.5	902.5	960.7	960.7	960.7	960.7	931.4	931.4	931.2	931.2
0.3	857.3	941.2	857.3	941.2	857.4	857.4	857.4	857.4	941.4	941.4	941.5	941.5	899.3	899.3	898.6	898.6
0.4	814.5	922.3	814.5	922.3	814.5	814.5	814.5	814.5	922.6	922.6	922.7	922.7	868.4	868.4	867.2	867.2
0.5	773.7	903.9	773.7	903.9	773.8	773.8	773.8	773.8	904.1	904.1	904.2	904.2	838.8	838.8	836.8	836.8
0.6	735.0	885.8	735.0	885.8	735.1	735.1	735.1	735.1	886.0	886.0	886.1	886.1	810.4	810.4	807.5	807.5
0.7	698.3	868.1	698.3	868.1	698.3	698.3	698.3	698.3	868.3	868.3	868.4	868.4	783.2	783.2	779.3	779.3
0.8	663.4	850.7	663.4	850.7	663.4	663.4	663.4	663.4	850.9	850.9	851.0	851.0	757.0	757.0	752.0	752.0
0.9	630.2	833.7	630.2	833.7	630.2	630.2	630.2	630.2	833.9	833.9	834.0	834.0	731.9	731.9	725.7	725.7
1.0	598.7	817.0	598.7	817.0	598.7	598.7	598.7	598.7	817.2	817.2	817.3	817.3	707.8	707.8	700.3	700.3
1.1	568.7	800.7	568.7	800.7	568.8	568.8	568.8	568.8	800.8	800.8	801.0	801.0	684.7	684.7	675.8	675.8
1.2	540.3	784.6	540.3	784.6	540.4	540.4	540.4	540.4	784.8	784.8	785.0	785.0	662.5	662.5	652.1	652.1
1.3	513.3	768.9	513.3	768.9	513.3	513.3	513.3	513.3	769.1	769.1	769.3	769.3	641.1	641.1	629.3	629.3
1.4	487.6	753.6	487.6	753.6	487.7	487.7	487.7	487.7	753.7	753.7	753.9	753.9	620.6	620.6	607.3	607.3
1.5	463.2	738.5	463.2	738.5	463.3	463.3	463.3	463.3	738.6	738.6	738.8	738.8	600.8	600.8	586.0	586.0
1.6	440.0	723.7	440.0	723.7	440.1	440.1	440.1	440.1	723.8	723.8	724.0	724.0	581.9	581.9	565.5	565.5

matching the X_1, Y_1 components with one set of weights and matching the X_2, Y_2 components with the other set.

Partial Aggregation and Disaggregation of Lanchester Systems

It is common to group systems by type in a combat model and create an aggregated weapon type for each such grouping. Fighter aircraft of similar capabilities might be grouped into equivalent fighter systems, or all tanks might be aggregated into an "equivalent" tank. This implies a *partial aggregation* of the system and, in terms of the earlier described aggregation mappings, we define R to have the canonical form

$$R = \begin{bmatrix} [\quad \rho_1 \quad][0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 0] \\ [0 \ldots\ldots 0][\quad \rho_2 \quad][0 \ldots\ldots\ldots\ldots\ldots 0] \\ \ldots \\ [0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 0][\quad \rho_r \quad] \end{bmatrix},$$

where each ρ_i is a nonnull vector of nonnegative weights on specific components of X. We rule out aggregation of a component into two different aggregate components, although it could be treated; but the treatment in what follows, particularly in the appendix, would be laborious in its details and not likely to be illuminating.

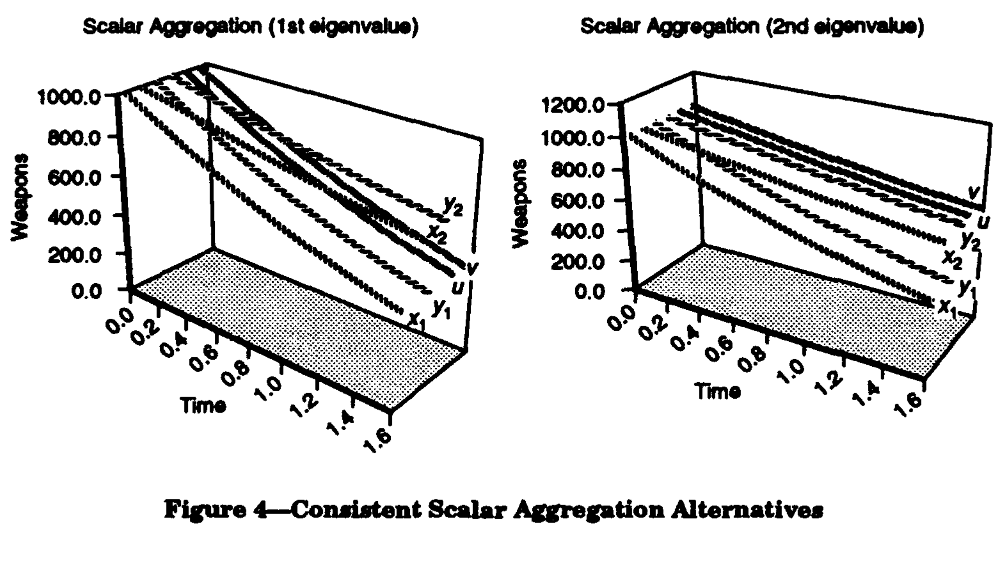

Figure 4—Consistent Scalar Aggregation Alternatives

For the heterogeneous Lanchester system, we can state the following two facts about the ability to do this and maintain consistency between the two models. These results are established in the appendix.

Fact 1 (Partial Aggregation). *One cannot generally partially aggregate a square-law Lanchester system so that* U *and* V *are vectors and consistency is maintained without applying additional restrictive conditions on the matrices* A *and* B.

Fact 2 (Disaggregation). *It is not generally possible to disaggregate a previously aggregated system and obtain a system consistent with the original disaggregated system without additional restrictive conditions on the matrices* A *and* B.

The condition required to partially aggregate or disaggregate a square-law Lanchester system is that there is a constant relative effectiveness or vulnerability between the various components of X and Y. For example, the following is shown in the appendix.

Fact 3 (Partial Aggregation). *Partial aggregation is possible (consistency maintained) when there is a constant proportionality in the effects of some of the weapons of a side with respect to all of the weapons of the other side.*

For example, if weapon 1 of side 1 is twice as effective as weapon 2 of the same side against all systems (components of Y), then weapon 1 and weapon 2 can be aggregated into a single representative weapon and this aggregated model will correctly predict the square-law attrition of each component of Y. This partially aggregated system model is consistent with the nonaggregated model if the weighting of the two weapons is proportional to their relative effectiveness.

Figure 5 illustrates this case, in which components x_1 and x_2 have been aggregated while x_3, y_1, y_2, and y_3 remain disaggregated. The trajectories of the components of Y remain the same regardless of whether the aggregated or disaggregated X is used in the equations.

This condition is not enough to be able to restore X from its aggregated state at some later time, however. This is because the system may have differential attrition of the components of X (x_1 and x_2 in the example) so that the number of surviving elements varies over time. Once aggregated, without additional information there is no way to turn U back into x_1 and x_2. What is a condition that would permit this? In the appendix we show the following is true.

Fact 4 (Disaggregation). *If, in addition to the previously stated relative effectiveness of two or more systems, the same systems also have a proportionality in vulnerability that is constant with respect to all systems of the opponent, then it is possible to aggregate these*

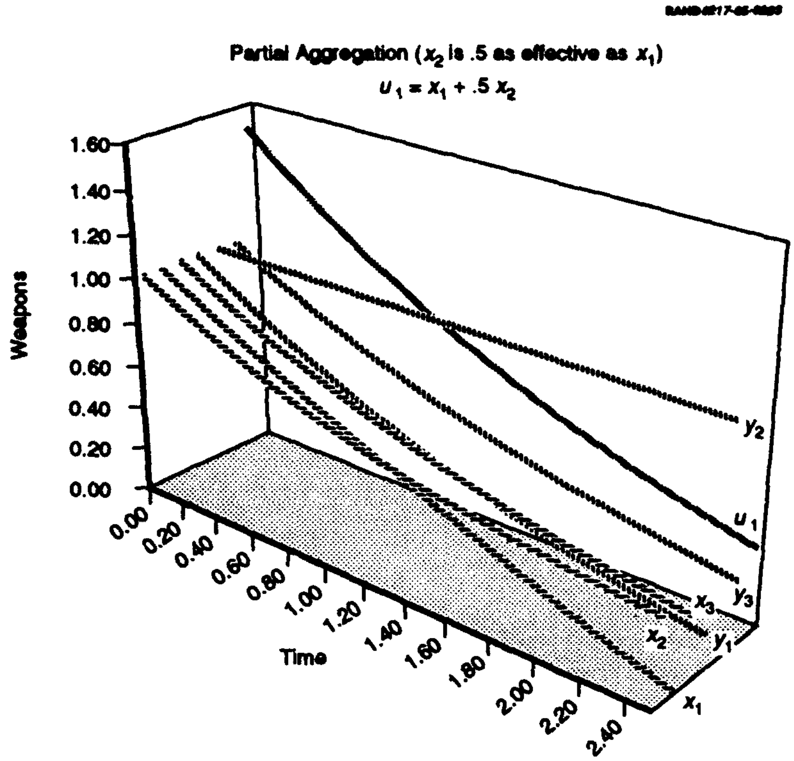

Figure 5—An Example of Partial Aggregation

systems consistently and at any point in time disaggregate the aggregated system back to the original more detailed model.

In other words, if we can say that weapon 1 is twice as effective against all weapons as weapon 2 and that weapon 1 is, say, one-third as vulnerable as weapon 2 to all opposing systems, then the detailed model can be partially aggregated and later disaggregated while satisfying the consistency requirement.

These are the conditions under which the scoring of weapons and later disaggregation of those scores make the most sense, since the aggregated values can be used without any loss of information. The information needed to restore the system is in the knowledge about relative effectiveness and vulnerability. As a final example, consider the three-component model shown in Table 2. The columns of the table show the consistency between the aggregation of the first two components of a three-dimensional example (three weapon systems) and the original system. The table also shows the reconstruction of the original system from the aggregated components (possible only because the respective components in the differential equations are linearly related). Note that

Table 2

Consistent Partial Aggregation and Disaggregation

A

	1	2	3
1	0.5	0.25	0.1
2	0.1	0.05	0
3	0	0	0.5

$U = (u_1, x_3)$

$dU/dt = -CV$

B

	1	2	3
1	0.4	0.2	0
2	0.12	0.06	0
3	0	0	0.5

$V = (v_1, y_3)$

$dV/dt = -DU$

C

	1	2
1	0.55	0.1
2	0	0.5

$u_1 = x_1 + p^* x_2$

D

	1	2
1	0.46	0
2	0	0.5

$v_1 = y_1 + r^* y_2$

	Original System						Aggregation Weighted		Aggregation Integrated		Aggregation Weighted		Aggregation Integrated		Disaggregation			
t	x_1	x_2	x_3	y_1	y_2	y_3	u_1	x_3	u_1	x_3	v_1	y_3	v_1	y_3	x_1d	x_2d	y_1d	y_2d
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.50	1.00	1.50	1.00	1.50	1.00	1.50	1.00	1.00	1.00	1.00	1.00
0.10	0.92	0.99	0.95	0.94	0.98	0.95	1.41	0.95	1.41	0.95	1.43	0.95	1.43	0.95	0.92	0.98	0.94	0.98
0.20	0.83	0.97	0.90	0.88	0.97	0.90	1.32	0.90	1.32	0.90	1.37	0.90	1.37	0.90	0.84	0.97	0.88	0.97
0.30	0.76	0.96	0.86	0.83	0.95	0.86	1.24	0.86	1.24	0.86	1.31	0.86	1.31	0.86	0.76	0.95	0.83	0.95
0.40	0.68	0.94	0.81	0.78	0.93	0.81	1.15	0.81	1.15	0.81	1.25	0.81	1.25	0.81	0.69	0.94	0.78	0.93
0.50	0.61	0.93	0.77	0.74	0.92	0.77	1.08	0.77	1.08	0.77	1.20	0.77	1.20	0.77	0.62	0.92	0.74	0.92
0.60	0.54	0.92	0.74	0.69	0.91	0.74	1.00	0.74	1.00	0.74	1.15	0.74	1.15	0.74	0.55	0.91	0.69	0.91
0.70	0.48	0.91	0.70	0.65	0.90	0.70	0.93	0.70	0.93	0.70	1.10	0.70	1.10	0.70	0.49	0.90	0.65	0.90
0.80	0.42	0.90	0.66	0.61	0.88	0.66	0.87	0.66	0.87	0.66	1.06	0.66	1.06	0.66	0.42	0.88	0.61	0.88
0.90	0.36	0.89	0.63	0.58	0.87	0.63	0.80	0.63	0.80	0.63	1.02	0.63	1.02	0.63	0.37	0.87	0.58	0.87
1.00	0.30	0.88	0.60	0.55	0.86	0.60	0.74	0.60	0.74	0.60	0.98	0.60	0.98	0.60	0.31	0.86	0.55	0.86
1.10	0.25	0.87	0.57	0.52	0.86	0.57	0.68	0.57	0.68	0.57	0.95	0.57	0.95	0.57	0.25	0.85	0.52	0.86
1.20	0.19	0.86	0.54	0.49	0.85	0.54	0.62	0.54	0.62	0.54	0.91	0.54	0.91	0.54	0.20	0.84	0.49	0.85
1.30	0.14	0.85	0.51	0.47	0.84	0.51	0.57	0.51	0.57	0.51	0.89	0.51	0.89	0.51	0.15	0.83	0.47	0.84
1.40	0.09	0.84	0.49	0.44	0.83	0.49	0.51	0.49	0.51	0.49	0.86	0.49	0.86	0.49	0.10	0.82	0.44	0.83
1.50	0.04	0.83	0.46	0.42	0.83	0.46	0.46	0.46	0.46	0.46	0.84	0.46	0.84	0.46	0.05	0.81	0.42	0.83
1.60	0.00	0.82	0.44	0.40	0.82	0.44	0.41	0.44	0.41	0.44	0.82	0.44	0.82	0.44	0.01	0.80	0.40	0.82

consistency is achieved until $t = 1.6$, at which time x_1 goes to zero. At this point the original Lanchester system must be reduced to leave out x_1 and all aggregations recalculated to maintain consistency. In this example we stopped the calculations at this point.

5. Conclusions

We have shown that aggregation and disaggregation should be performed carefully in models of combat if consistency is to be preserved; but it can be done, provided that the attrition matrices satisfy certain additional reasonable conditions when the aggregations are not to be total. Although we do not suggest that the Lanchester square law is a realistic depiction of actual conflict, the fact that, even for this linear system of equations, consistent aggregation and disaggregation cannot be done without severe restrictions on the attrition matrices A and B implies that ad hoc approaches to varying resolution may not lead to consistent models. The aggregation weights cannot be arbitrary. Once the attrition matrices A and B are specified, the aggregation weights, when consistent aggregation is possible, are dictated by mathematical considerations, as we show in the appendix.

Clearly, the absence of general conditions of the type shown here for more complex models implies that empirical tests comparing models of different resolution should be made before conclusions can be drawn regarding the goodness of an aggregation.

We do not mean to argue that detailed models are always better. Often the aggregate results of combat (overall losses and advances of front lines) are known from history, but the details of specific forces, conditions, tactics, effectiveness, fire allocation, and so forth are missing. This means that an aggregate model can be tested or fit to the data, but an attempt to extrapolate to detailed losses is highly subjective. Thus, the most correct model based on empirical data could very well be the aggregate, low-resolution one. On the other hand, the frequent absence of any empirical data on how forces or weapons might fare in battle has often forced analysts to build models in high detail in hopes that engineering test data can be extrapolated to combat outcomes. Frequently, however, this approach amounts to compounding assumption upon assumption regarding interactions in conflict—assumptions that are completely subjective.

The material in this research sheds very little light on this problem. It does suggest, as noted earlier, that aggregation and disaggregation cannot be done arbitrarily and that fairly strong requirements are necessary to obtain consistent high- and low-resolution models.

Further research in this area should examine when models and aggregations can be made partially consistent. That is, suppose the conditions for aggregations and disaggregations stated in this report are approximately satisfied. What does this imply for how far apart the aggregate solution of the unaggregated differential equations is from the solution of the aggregated differential equations as time advances?

Another area of investigation to consider is the possibility of extending this research to Lanchester models with nonlinear differential equation formulations that may be equivalent to higher-order linear ones as, for example, certain higher-dimensional Riccati differential equation systems with special structure are.

Appendix
MATHEMATICAL CONSIDERATIONS

Preliminaries

We consider here a rather sharply defined Lanchester system model of continuous combat between two opposing forces whose resources fall into different classes or types, e.g., tanks of one or more types, missile launchers, personnel vehicles, howitzers, helicopters, troops of different types, etc. The numbers of each resource type form the different components of a strictly positive $m \times 1$ vector X for one side and of a strictly positive $n \times 1$ vector Y for the other side.

The model assumes that the state of the battle at a time $t > 0$ is represented by the pair (X, Y), which are obtained as the solutions of the system of ordinary equations mentioned earlier:

$$\begin{aligned} \frac{dX}{dt} &= -AY, \quad X(0) = X_0 > 0 \\ \frac{dY}{dt} &= -BX, \quad Y(0) = Y_0 > 0 \ , \end{aligned} \tag{1}$$

where $A: m \times n$ and $B: n \times m$ are the nonnegative attrition matrices. The elements a_{ij} of A represent the time rate of attrition of the ith resource x_i by one unit of the jth resource y_j of the opponent. The elements b_{ij} of B are similarly defined with x and y interchanged.

We note that the components of the vectors, X and Y, in the solution of equation (1) are monotone-decreasing functions of time. When any component, or components, of these vectors reaches zero, we consider the time duration of this system, but not necessarily the battle, to be terminated. A new system, if the battle is to continue, is created, with its dimensions reduced by the number of components that have reached zero; these components and those rows and columns of the matrices that are associated with them are removed.

The new system is still a Lanchester system with the same generic form and properties the system (1) has, with initial conditions for the positive vectors being the respective values they attained at the end of the previous time period.

Although the same battle continues, we simply consider it as continuing under a new system, new only in the sense that its dimensions have been reduced and its initial conditions changed.

This process of "culling" the system of variables and associated rows and columns in the attrition matrices (that could, if not culled, lead to the absurdity of negative numbers of resources) can be continued so long as there are positive components left on each side, or stopped earlier if it is deemed that the battle is over.

We further make a reasonable assumption of choice that each resource on either side has some attriting effect on at least one resource of the opponent, and each resource on either side is vulnerable to attriting effects of at least one resource on the opponent's side. Thus, not only are A and B matrices of nonnegative components, we require also that they have no completely null column and no completely null row. This property of no null columns and no null rows in the attrition matrices, with their nonnegativity, is sufficient for their respective products to satisfy the hypotheses of one or the other of the Perron-Frobenius theorems on the existence of a maximal positive eigenvalue and corresponding positive or nonnegative eigenvector for nonnegative matrices that we will use later on.[1] (Of course, one can easily conceive of resources with value that is not of an attriting nature, e.g., petroleum supplies, or with an invulnerable nature, e.g., aircraft against an enemy with no air defenses; and one can deal with such cases. However, for some mathematical convenience later, namely, to avoid the potential for zero Perron eigenvalues and concomitant complications, we exclude these cases by our assumptions above.)

As is to be expected from the general theory for systems such as (1), its solution is an analytic vector function of the initial condition vectors. We define

$$F(z) = \sum_{k=0}^{\infty} \frac{t^{2k} z^k}{(2k)!} \quad \text{and} \quad G(z) = \sum_{k=0}^{\infty} \frac{t^{2k+1} z^k}{(2k+1)!} . \tag{2}$$

Then, we have for the solution of (1)

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} F(AB) & -G(AB)A \\ -G(BA)B & F(BA) \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} , \tag{3}$$

---

[1]See, e.g., chapter 2 of R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall Inc., Englewood Cliffs, N.J., 1962.

for as long as no component of the solution vector becomes negative. For $z = AB$ or for $z = BA$, computing a solution in the form (3) can be laborious; consequently, employing the similarity transformations,

$$AB = W\Lambda W^{-1} \quad \text{and} \quad BA = Z\Gamma Z^{-1} ,$$

we have a much more computationally tractable form:

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} WF(\Lambda)W^{-1} & -WG(\Lambda)W^{-1}A \\ -ZG(\Gamma)Z^{-1}B & ZF(\Gamma)Z^{-1} \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix},$$

where the diagonals of the Jordan canonical form matrices (even possibly diagonal), Λ and Γ, contain the eigenvalues of AB and BA, respectively.

However, it may be that the dimensionality of the problem is still too high, either for computational reasons or, more importantly, because the detail is too fine for understanding the progress of the battle at various levels of generality or resolution. Thus, to relieve these objections to the size of the dimensionality of the original problem, one frequently employs weighted linear aggregations of the resources on either side, either into a single scalar value for each side (total aggregation) or into smaller numbers of groups of resources than the original numbers (partial aggregation) or, perhaps, even total scalar aggregation on one side and no aggregation on the other (unilateral aggregation). Having presented some illustrative examples earlier, our purpose here is to examine mathematically some aspects of aggregation and conditions for potential disaggregation.

We define linear dimension-reducing aggregation operators, R and S, to be $r \times m$ and $s \times n$ nonnegative matrices, respectively, with r and s strictly less than m and n, respectively, such that

$$U = RX \quad \text{and} \quad V = SY \tag{4}$$

form a reduced-dimension Lanchester system in the above-defined sense. Thus, we have a system of ordinary differential equations for U and V of similar form to (1) for X and Y:

$$\frac{dU(t)}{dt} = -CV(t),\ U(0) = U_0 > 0 ,$$

$$\frac{dV(t)}{dt} = -DU(t),\ V(0) = V_0 > 0\ , \tag{5}$$

where C and D are new $r \times s$ and $s \times r$ nonnegative attrition matrices, respectively, with the same properties of having nonnull rows and nonnull columns as A and B.

The solution to the reduced-dimension aggregated Lanchester system (5) is of exactly the same form as the solution (3) of (1), namely,

$$\begin{bmatrix} U(t) \\ V(t) \end{bmatrix} = \begin{bmatrix} F(CD) & -G(CD)C \\ -G(DC)D & F(DC) \end{bmatrix} \begin{bmatrix} U_0 \\ V_0 \end{bmatrix} . \tag{6}$$

The aggregation operators R and S appearing in (4) are defined to be in canonical form if they are in the forms illustrated below:

$$R = \begin{bmatrix} [\quad \rho_1 \quad]\,[0 \ldots\ldots\ldots\ldots\ldots\ldots 0] \\ [0 \ldots\ldots 0]\,[\quad \rho_2 \quad]\,[0 \ldots\ldots\ldots\ldots 0] \\ \cdots \\ [0 \ldots\ldots\ldots\ldots\ldots\ldots 0]\,[\quad \rho_r \quad] \end{bmatrix} \tag{7}$$

and

$$S = \begin{bmatrix} [\quad \sigma_1 \quad]\,[0 \ldots\ldots\ldots\ldots\ldots\ldots 0] \\ [0 \ldots\ldots 0]\,[\quad \sigma_2 \quad]\,[0 \ldots\ldots\ldots\ldots 0] \\ \cdots \\ [0 \ldots\ldots\ldots\ldots\ldots\ldots 0]\,[\quad \sigma_s \quad] \end{bmatrix} , \tag{8}$$

where the row vectors $\rho_i, i = 1,\ldots,r$ and $\sigma_i, i = i,\ldots,s$ are nonnull and nonnegative and are of respective dimensions m_i and n_i. For convenience, we shall consider aggregation operators in the above canonical forms (7) and (8). (Conceivably one could presume aggregations that allow a resource to be in two or more aggregate groups; but because of the additional complexity of details in their treatment, we conveniently refrain from examining such cases.)

A central desideratum for aggregation operators, one that makes for a sense of consistency, is that the operations of differential equation solving and aggregation commute. That is, the result obtained by first solving the ordinary differential equation system (1) to obtain the solutions given by (3), or of any other equivalent form, and then applying the dimension-reducing aggregation operators, *R* and *S*, to the solutions is identical to the result of first applying the operators, *R* and *S*, to the differential equations (1) to get the differential equation system (5) in *U* and *V* and then solving this system (5) to get the reduced-dimension result (6).

In the following sections we shall be concerned with presenting conditions on *A*, *B*, *R*, *S*, *C*, and *D* that will produce the desired commutativity in the different situations of total (scalar) and partial aggregation; verification that the matrices *AB* and *BA* satisfy the hypotheses of one or the other of the Perron-Frobenius theorems mentioned below, which are useful in determining *R* and *S* that preserve the desired commutativity; and conditions for the inversion of aggregation, i.e., disaggregation.

General Conditions for the Preservation of Lanchester System Property and Commutativity of Aggregation and Differential Equation Solving

Theorem 1. *If aggregation operators* R, r × m, *and* S, s × n, *and nonnegative matrices* C *and* D *of compatible dimensions exist such that*

$$RA = CS \qquad \textit{and} \qquad SB = DR, \tag{9}$$

then U *and* V *defined by (4) will produce a Lanchester system (5) from the Lanchester system (1). Moreover, with compatible initial conditions, the solution (6) of (4) will be identical to the pair,* (RX, SY), *where* (X, Y) *is the solution (3) of (1).*

Proof. Let X and Y satisfy a Lanchester system of differential equations (1). Suppose that nonnegative *R*, *S*, *C*, and *D* exist such that (9) holds. Then we have

$$\frac{dU}{dt} = R\frac{dX}{dt} = -RAY = -CSY = -CV$$

and

$$\frac{dV}{dt} = S\frac{dY}{dt} = -SBX = -DRX = -DV\,.$$

Hence, with positive initial values, U_0 and V_0, for U and V, respectively, U and V form a Lanchester system pair as in (5).

Now, consider any analytic function, $f(z)$ (particularly $F(z)$ and $G(z)$ defined in (2)), representable by an absolutely convergent power series in z in some disk of the complex plane with a radius exceeding the spectral radii of AB, BA, CD, and DC (those of AB and BA are equal, as are those of CD and DC). Then, applying the associative law for matrices successively, we have that the relations (9) imply that $R(AB) = C(SB) = (CD)R$, $S(BA) = (DC)S$, $R(AB)^k = (CD)^k R$, $S(BA)^k = (DC)^k S$, and, finally,

$$Rf(AB) = f(CD)R \qquad \text{and} \qquad Sf(BA) = f(DC)S \ . \tag{10}$$

Let the U and V satisfy (5) with initial values of $U_0 = RX_0$ and $V_0 = SY_0$, respectively. Then, applying R to X, S to Y, and (10) with f replaced by F and G as defined in (2) in the solution (3) to equation (1) completes the proof of the theorem. *Q.E.D.*

Bilateral Total (Scalar) Aggregation

In this section, we are concerned with defining the row vectors R and S which totally aggregate X and Y down to scalars and, consequently, make U, V, C, and D positive scalars. We will make use here of the two Perron-Frobenius theorems:[2]

(i) *Let* M *be a nonnegative, irreducible square matrix. Then it has a positive eigenvalue (called the Perron eigenvalue) that is simple and is not exceeded by the absolute value of any of its other eigenvalues. Furthermore, corresponding to this eigenvalue is a positive eigenvector (left eigenrow or right eigencolumn).*

(ii) *Let* M *be a nonnegative, reducible, square matrix, but for which there exists no permutation matrix,* P, *such that* PMP^T *is a strictly upper triangular matrix. (The superscript* T *denotes transpose.) Then it has a positive eigenvalue that is equal to its spectral radius and a corresponding nonnegative, nonnull left eigenrow and right eigencolumn.*

(A square matrix, M, is reducible if there exists a permutation matrix, P, (i.e., a square matrix whose only nonzero elements are a single 1 in each row and in each column) such that

---

[2] Ibid.

$$PMP^T = \begin{bmatrix} N_1 & N_2 \\ 0 & N_3 \end{bmatrix},$$

where N_1 and N_3 are square matrices and 0 is a null matrix; M is irreducible if no such permutation matrix exists. A square matrix is strictly upper triangular if all of its elements are zero except those strictly above the main diagonal, which are unrestricted.)

Theorem 2. *If no row or column of either of the nonnegative matrices* A *or* B *is a null vector, then there exist no permutations,* P *or* Q*, such that either* $AB = PABP^T$ *or* $BA = QBAQ^T$ *is similar to a strictly upper triangular matrix.*

(Hence, AB and BA, being nonnegative, also satisfy the hypotheses of one or the other of the Perron-Frobenius theorems, depending on whether they are irreducible or reducible.)

Proof. For the sake of a contradiction later, let us assume that, by permuting the rows and columns of AB (via the same permutation), AB is transformed into a strictly upper triangular matrix. Then this strictly triangular matrix has a null first column and a null last row. Let the column of AB that becomes the first column under the permutation be the jth column for a particular j. Let K be that set of integers k such that $b_{kj} > 0$ whenever $i = k$. K is not empty, by the hypothesis of the theorem. Since $a_{ik}b_{kj}$ must equal zero for every $i = 1,...,m$ and $k = 1,...,m$ (AB being $m \times m$), the kth columns of A must be null whenever k is in K, contrary to the hypotheses of the theorem.

Similarly, let the row of AB that becomes the last under the permutation be the ith for a particular i. Anew, let K be that set of integers k such that $a_{ij} > 0$ whenever $j = k$. Arguing similarly to the above, we must have that the kth rows of B must be null whenever k is in K. Again, we have a contradiction to the hypotheses of the theorem.

Reversing the roles of A and B in the arguments above and applying them to the $n \times n$ matrix BA, we again arrive at the findings that A must have some null row or rows and B must have some null column or columns, contrary to the hypotheses of the theorem. Therefore, neither AB nor BA is similar under permutation to a strictly upper triangular matrix. *Q.E.D.*

Thus AB and BA satisfy the hypotheses of the Perron-Frobenius theorems. Now it is easy to demonstrate that if A and B are square matrices, then they have the same set of eigenvalues; but what can be said if they are not?

Let λ and ρ be the Perron eigenvalue and corresponding left eigenrow for AB, and μ and σ be the same for BA. For convenience, though it is not necessary, let the norms of these two eigenrows be unity.

Lemma 1. *With the above definitions, we have that* $\lambda = \mu$ *and, if these eigenvalues are simple, then* $\rho A \,/\, |\rho A| = \sigma$ *and* $\sigma B \,/\, |\sigma B| = \rho$.

Proof. The assumption that λ and μ are Perron eigenvalues of the respective matrices AB and BA implies that λ is also an eigenvalue of BA and therefore $\lambda \le \mu$. Similarly, μ is also an eigenvalue of AB and therefore $\mu \le \lambda$. Hence, $\lambda = \mu$.

If AB and BA are irreducible, then their Perron eigenvalue is simple. If they are reducible, then we must assume additionally that the maximum of all the Perron eigenvalues of the submatrices on the main diagonal is unique. Then, multiplying $\rho(AB) = \lambda\rho$ by A on the right on both sides and applying the associative law, we have that $(\rho A)(BA) = \lambda(\rho A)$. Hence, since λ is simple, we have that $\rho A \,/\, |\rho A| = \sigma$. Similar argument yields that $\sigma B \,/\, |\sigma B| = \rho$. *Q.E.D.*

Therefore, for total aggregation on both sides, we define

$$R = k\rho \,/\, |\rho A| \qquad \text{and} \qquad S = \kappa\sigma \,/\, |\sigma B| \;.$$

In order to achieve our desideratum of consistency, i.e., commutativity of aggregation and differential equation solving, we impose the hypothesis of Theorem 1 and the requirement that the Perron eigenvalues of AB and BA are simple if so needed when these matrices are reducible. Then, from the above lemma, we have that

$$RA = \frac{k(\rho A)}{|\rho A|} = CS = \frac{\kappa C\sigma}{|\sigma B|} = \frac{\kappa C(\rho A)}{|\sigma B|\,|\rho A|}$$

and

$$SB = \frac{\kappa(\sigma B)}{|\sigma B|} = DR = \frac{kD\rho}{|\rho A|} = \frac{kD(\sigma B)}{|\rho A|\,|\sigma B|}\;,$$

from which we get

$$k = \frac{\kappa C}{|\sigma B|} ,$$

$$\kappa = \frac{kD}{|\rho A|} ,$$

and

$$CD = |\rho A|\,|\sigma B| .$$

Furthermore, from this, $\rho AB = \lambda\rho, \sigma BA = \lambda\sigma$, and the lemma, when the eigenvalues are simple, we have that $\lambda\rho = \sigma B|\rho A|$ and $\lambda\sigma = \rho A|\sigma B|$. Taking absolute values of either, we conclude that $\lambda = |\rho A||\sigma B|$ and $CD = \lambda$. Some freedom of scaling in aggregation exists; but C, D, k, and κ are not entirely arbitrary. They must satisfy the above relations.

This concludes the definition of the total (scalar) aggregation operators (row vectors, in this case) sufficient to produce commutativity of the aggregation and the solution of the Lanchester differential equations.

Unilateral Total Aggregation

For the sake of definiteness, let us assume that it is the resources in the vector X that are to be aggregated into a single somehow representative resource, $U = RX$, and that the vector Y is not to be aggregated; i.e., the aggregation operator S is replaced by the identity operator, and $V = Y$. At the outset, we assume the aggregation row vector R is determined from considerations external to the mathematical ones above. Of course, we should have the desired previously mentioned commutativity. This is not possible with arbitrary A and B as it was in the case of bilateral total aggregation, as we shall see.

The relations $U = RX$ and $V = Y$, the equations (5) with U being a scalar, and

$$\frac{dV}{dt} = -BX = -DU = -DRX$$

imply that every row of B must be proportional to R where the constants of proportionality are the elements of D. We note further that if all the rows of a matrix are proportional to some common row, here R, then also all the columns must be proportional to some common column, here D, the constants of proportionality being the elements of R. That is, the elements of B must be of the form $p_i q_j$.

Conversely, if the elements of B are of this form, then to preserve the desired commutativity, R must be proportional to a row vector whose elements are q_j, $j = 1,...,m$. Consequently, we have proved the following:

Theorem 3. *Unilateral aggregation of* X *with no aggregation of* Y *is possible if, and only if, the elements of* B *are of the form* $p_i q_j$, $i = 1,...,n$, $j = 1,...,m$, *and then the aggregation vector,* R, *is proportional to any row of* B.

The attrition operator C is then determined from $RA = C$.

Clearly, interchanging X and Y and m and n, and replacing B by A and R by S, we have the same unilateral aggregation theorem for the opponent's resource vector.

Incidentally, we note that the above form for B is not unreasonable from the point of view of modeling. It essentially postulates an attrition matrix element b_{ij} against Y due to X, which is the product of a generalized overall "average" vulnerability factor p_i for the resource Y_i and a generalized overall "average" lethality factor q_j for the resource X_j.

Unilateral Partial Aggregation

Again for definiteness, we consider aggregation on the resource vector X with the aggregation operator, R, where the dimension r exceeds unity and with no aggregation on Y. Here we have that $U = RX$ and $V = Y$. Consequently, $RA = C$ and $B = DR$ are the sufficient conditions to achieve the desired commutativity and Lanchester consistency. Since D here is of dimension $n \times r$ and R is of dimension $r \times m$, it is clear that it is necessary that there be blocks of columns in B whose numbers of columns correspond to the lengths of the component subvectors in the canonical representation of the aggregation operator R shown in (7).

If we assume that a permutation of the indices of the resources is made so that such blocks become blocks of contiguous columns, a block for each aggregation subvector in the canonical form of R, we then see that these blocks in B have the same form $[p_i q_j]$ as B had in the previous section on unilateral total aggregation but over a restricted integer interval of values of j for each block. Then the component subvectors in R become row vectors proportional to the particular row vectors $[q_j]$ that are common to the pertinent block of columns in B. And the corresponding columns of D become column vectors proportional to the pertinent common column $[p_i]$, their constants of proportionality being the reciprocals of the constants of proportionality in the subvectors of R. Thus R and D are determined up to scale constants that are reciprocally related.

The attrition matrix, C, on the right-hand side of the equation for dU/dt for the reduced Lanchester system (5) is obtained directly from $RA = C$. Once again, as in the subsection above, we arrive at necessary and sufficient conditions, $RA = C$ and $B = DR$, with a particular structure for B, that permit a unilateral aggregation and either define R or, if R is determined exogeneously, define rows of B. B cannot be arbitrary. If it is arbitrarily chosen, aggregations with the desired consistency and commutativity properties cannot be achieved.

Bilateral Partial Aggregation

Here we revert to Theorem 1 giving sufficient conditions on the relevant matrices, A, B, R, S, C, and D, that provide the desired commutativity and consistency—that is, of course, when A and B have the appropriate structure. In other words, we require that $RA = CS$ and $SB = DR$.

The arguments are similar to those in the preceding subsection and will not be carried out in detail. However, there are a few points that are different or that need some minor modification.

We have seen in the preceding subsection that the prior specification of the lengths of the blocks of resources in X that are to be aggregated determines the numbers of columns in B that are proportional to a common column as well as the common column up to a multiplicative constant or, conversely, that the appearance of sets of columns in B that are proportional to a common column determines the aggregation subvectors in R, again up to a multiplicative scale factor.

We also note that if there are sets of columns of B or of A that are proportional to a common column, then SB and RA respectively have the same respective sets of columns that are proportional to a common column (not the identical columns, but ones that are those columns multiplied on the left by S or R respectively). Consequently, again R will be composed of subvectors that will be proportional to row vectors whose components will be those constants of proportionality in each block in B sharing a common column vector. Conversely, if R is specified, then the column indices of blocks of columns in B that have to be proportional to a common column are identified and, up to a multiplicative scalar factor, the row vectors of the constants of proportionality in these blocks are proportional to the corresponding components in *R*.

Similarly, corresponding properties of *A* are determined by a priorly determined *S*, and conversely, properties of *S* by a priorly determined *A*. The attrition matrices *C* and *D* are then determined by solving the linear equations $RA = CS$ and $S\cdot = DR$ for the elements of *C* and *D* respectively. This should always be poss:' to do uniquely if the full aggregation possibilities afforded by the blocks of proportional columns in *A* and in *B* are effected. If this is not done, then we have undetermined systems and degrees of freedom because choices in *C* and *D* exist.

Disaggregation

Obviously, disaggregation to the unaggregated *X* and *Y* directly from *U* and *V* in the solution (6) of the system of aggregated differential equations (5) without using other information is impossible, no more possible than, for example, determining the value of two numbers from their arithmetic average without another piece of information, such as their difference, another differently weighted average, or some other functional relationship between the two.

In the case of completely general *A* and *B* that, of course, still satisfy our original condition of not possessing any null columns or rows, there are two possible options (whose full description is outside the intent of the current work) to provide the additional information. They are both based on solving the eigenvalue problems for *AB* and *BA* (really only one eigenvalue problem, if *A* and *B* are square).

Omitting complicating details that can occur when either or both of the matrix products *AB* and *BA* are not diagonalizable, we know that the solution (*X*, *Y*) of the original system (1) can be written in terms of two linear combinations of exponentials involving square roots of the eigenvalues of *AB* and *BA* whose coefficients can be determined from the solution (*U*, *V*) of (5) in either of two ways.

The solution (*U*, *V*) of the reduced Lanchester system (5) can also be considered in terms of linear combinations of the very same exponentials, although some of the coefficients are zero because the system is reduced.

The unknown coefficients for two linear combinations of exponentials are related through two underdetermined systems of linear equations to the now known coefficients in the linear combination coefficient representations of *U* and *V*, where the two matrices defining the two systems have the exponentials evaluated at a given time, say, the current. The systems can now be determined by either storing the solution (*U*, *V*) of (5) for a succession of values of time sufficient to make the two systems of linear equations in the unknown coefficients determined, or by differentiating *U* and *V* a sufficient number of times for the current or any other value of *t* to enlarge the two systems of equations in the unknown coefficients to become determinate.

Now, for the not-so-general cases, suppose that *X* is either totally aggregated or partially so. Now suppose that, in addition, all rows of *A*, in the case of total aggregation, are proportional either to some common row or to blocks of rows of *A* coinciding with the blocks of components of *X* that are being aggregated into a single component of *U*, in the case of partial aggregation.

Then, it is essentially obvious that each component of *X* in the totally aggregated *X* case is linearly related to the solution *U*, the vector of slopes of these linear relations being proportional to the vector of constants of proportionality among the rows of *A*. The intercepts are determined from the differences between the initial conditions for *U* and the initial conditions for the particular component of *X* of immediate concern. The arguments in the case of partial aggregation are identical, except that they are applied individually to each aggregated block of components of *X* and the associated component of *U*. Of course, the same arguments are applied to any aggregation, total or partial, of *Y*, provided that all or else the relevant blocks of rows of *B* are proportional to common rows respectively.