FITTING KOREAN WAR DATA BY STATISTICAL METHODS

by
John L. Overholt

Professional Paper No. 32.

CENTER FOR NAVAL ANALYSES

May 5, 1970

15

PROFESSIONAL PAPER NO. 32

Fitting Korean War Data By Statistical Methods

This paper was delivered at the National Gaming Council panel on validation of games, Washington, D. C. April 19, 1970.

by

John L. Overholt

FITTING KOREAN WAR DATA BY STATISTICAL METHODS

John L. Overholt
Center for Naval Analyses

Abstract

Korean war data obtained from the records of all services is related to U.N. divisions engaged with North Korean and Chinese Communist troops in central Korea in May 1951. All data was taken from U.N. force records, which contained estimates of enemy strengths, casualties, and weapons. The daily record for each division was considered: the strengths and casualties of opposing forces, the amount of ground gained or lost, and fire support as measured in artillery rounds and air sorties.

Various graphs display changes in individual variables and in combinations of variables, from day to day and with undulations in the battle line as forces concentrated to attack or withdrew. These plots suggested more sophisticated treatment, with the use of multiple regression to fit various proposed relationships. These were relationships between casulaties and probable causes, such as the quantity of ordnance fired, the intensity of the battle, and the exposure of the troops as indicated by ground gained or lost. The two major problems were missing data and the correlation between types of data that probably had a common cause that could not be measured directly, such as the decisions of the opposing division (or higher) commanders to advance, hold, or withdraw, and their willingness to accept casualties to achieve objectives.

Lanchester equations of various types were examined by the incorporation of firepower into the strength and casualty relationships. Some empirical equations were also prepared.

The data and results were presented briefly; this current work will be presented elsewhere in an extended form at a later date.

Presentation

Ability to predict the outcome of battles would be useful for generals, military historians, and analysts. Lanchester proposed relationships between strengths and casualties; though these relationships are often used in war gaming and in studies, they have not been verified successfully very often. To our knowledge, no one has, for instance, analyzed the changes that occur in a continuing battle.

Therefore, as part of a Center for Naval Analyses study that modeled the effect of fire support on ground warfare, we examined the records of the U.N. divisions and corps engaged in central Korea in May 1951. The Air Force, Navy, and Marine Corps records were also consulted to supply data that was not available in Army records.

The daily situation maps (figure 1) provide front length and ground gained. To obtain the numbers of air sorties, we compiled the impact points of ordnance and related them to the units supported. This data (as in figure 2) provides an index to the intensity of action over the entire front when Army records are missing.

Many plots are made to show the continuity of action from day to day. The movement of divisions throughout the month is shown in figure 3. The front was quiet during the first half of the month, with small patrols engaging from time to time. The Chinese Communists massed forces in front of the 5th Republic of Korean (ROK) division, achieving a 1.4-to-1 force ratio at the start of the attack on 15 May.

The attack came in massed human waves which attempted to overrun the U.N. positions; the enemy was armed with smaller weapons only and had no artillery or air support. The U.N. forces planned to inflict maximum casualties with minimum losses to themselves; ground gains were not regarded as important. The ROK unit gave way, and the adjacent divisions were forced back to prevent a breakthrough. During this period the U.N. forces inflicted terrific losses: on 38 division-days, enemy losses exceeded 1000 per division; at times, losses per division mounted to 8000 or more per day. The median exchange rate in major attacks between enemy and U.N. forces was 70-to-1 during enemy offensives and 35 to 1 during our counterattacks in the latter part of the month.

The cumulative casualties were plotted on log-log paper to condense results. This plot (figure 4) shows the units that took heavy casualties during the enemy's attack and our counter-attack. The data was eventually reduced to a record of what happened per mile of front; the support was converted from rounds of artillery of various calibers, light and heavy aircraft sorties, and the estimated strengths of each side, into tons of ordnance fired. The tonnage of artillery and air ordnance per mile of front for each division (shown in figure 5) was greatest during the enemy attack; usage diminished as the U.N. forces attacked.

Multiple regression analysis was used to find the statistically significant terms in various relationships. There are two difficulties: Data was sometimes missing or bad, especially when our forces were on the defensive and falling back, and some units kept poor records or lost them in battle.

During battles, there is an increase in nearly all the factors measured. Thus, there is a high correlation among the various kinds of fire support, casualties, and ground gained. It is often difficult to decide whether any cause-and-effect relationship should be inferred; it is easier to decide that some relationships are obviously useless, such as the high correlation between friendly casualties and number of light artillery rounds fired. Relationships such as these reflect a common cause, which often was the commander's decision to attack.

Five forms of the Lanchester equations are tested in this analysis. Enemy casualties are predicted quite well; friendly casualties are not. The former is based on data on our fire-power and strength, which we know well; our casualties are predicted entirely from enemy strength, of which our knowledge is rather poor. None of the Lanchester equations proves markedly superior to another when the exponent to strength of force equals one. Some of the empirical equations in which the exponents are fitted indicate that the log law holds when the enemy attacked in strength; for other purposes the square law is as good as any other.

The conclusions about the equations are based on whether the coefficients of the terms were statistically significant, at the 95 percent level. The percentage of variation in the data explained by the entire equation is the regression coefficient R^2. If the value of R^2 is rather high, changing the form of the equation does not usually change the significant terms in the equation. When R^2 is low, the conclusions are often inconsistent when the model is changed.

Typical R^2 results are displayed in figure 6. The data is separated into 3 types of battle. The friendly casualty equations all have low R^2 values; in these equations, enemy strength is not important. Generally, enemy troops are killed by heavy fire before they engage the U.N. forces. Divisional strength data seems to include many forces that are not engaged; perhaps better results would be obtained by consideration of only the maneuver elements engaged. This requires examination of smaller units and would take much more time. During patrol periods, the enemy strength is not important, either, and enemy casualties are related to our firepower. The two lines labeled "separate" and "total" refer to separate forms of ordnance and total ordnance tonnage. Although, there is usually not much difference in R^2 value between the two, the former indicates the kind of ordnance that is more important.

In the last figure, a few equations are presented covering enemy casualties under 4 battle conditions. During heavy battles, our light and heavy bombers caused most enemy casualties when the U.N. was advancing; when the enemy attacked, however, our light artillery was most important. During patrol periods, our artillery caused equal numbers of deaths, no matter which side was attacking.

This, then, is a resume of the study that is now being prepared for publication. We hope that the work will be available with the data so that others can try their hand. Because of gaps in the data, we were unable to include all the important factors. The study has provided insight into the weaknesses of attempts to model wars on the basis of divisional strengths alone.

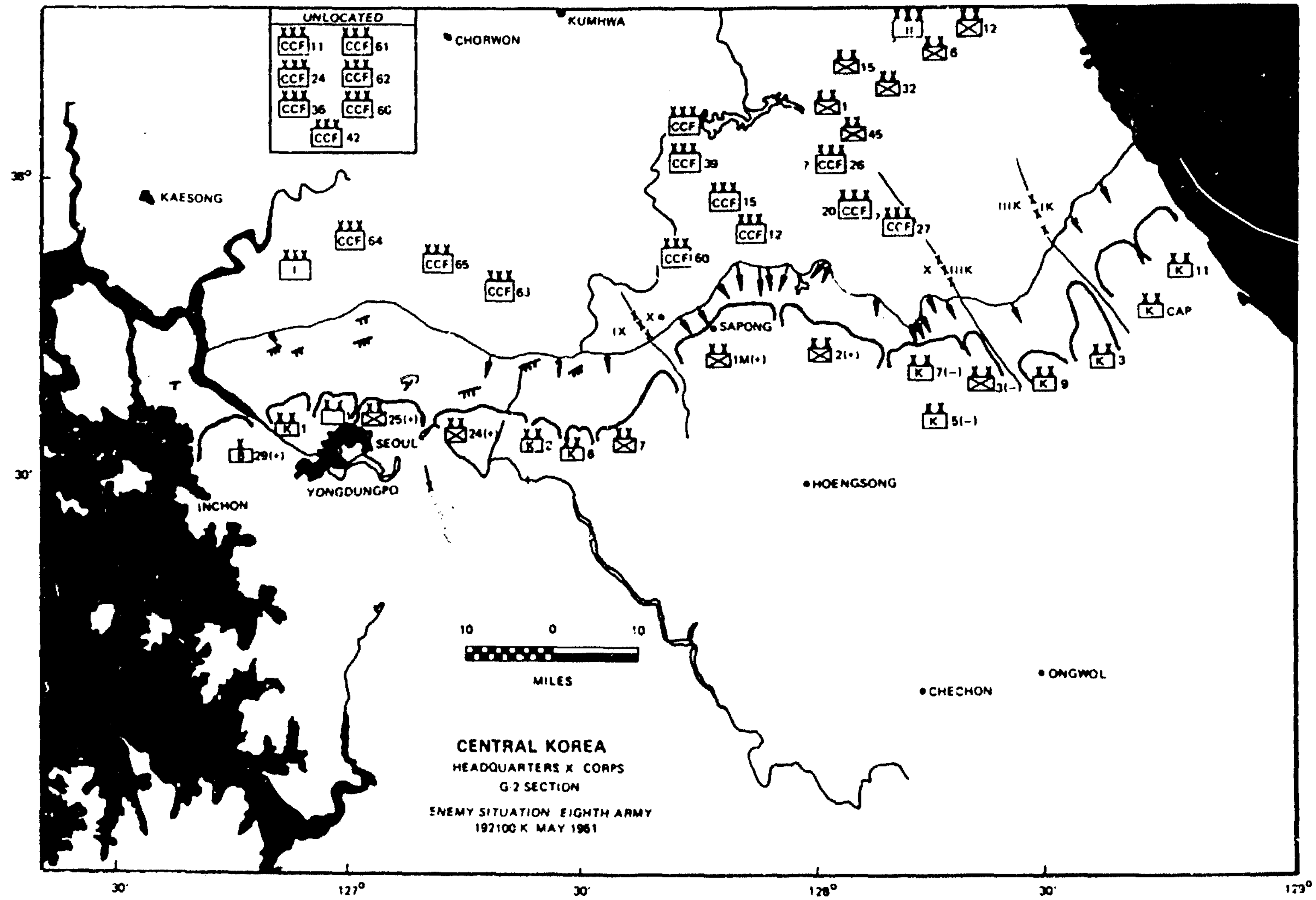

FIG. 1: SAMPLE DAILY SITUATION MAP

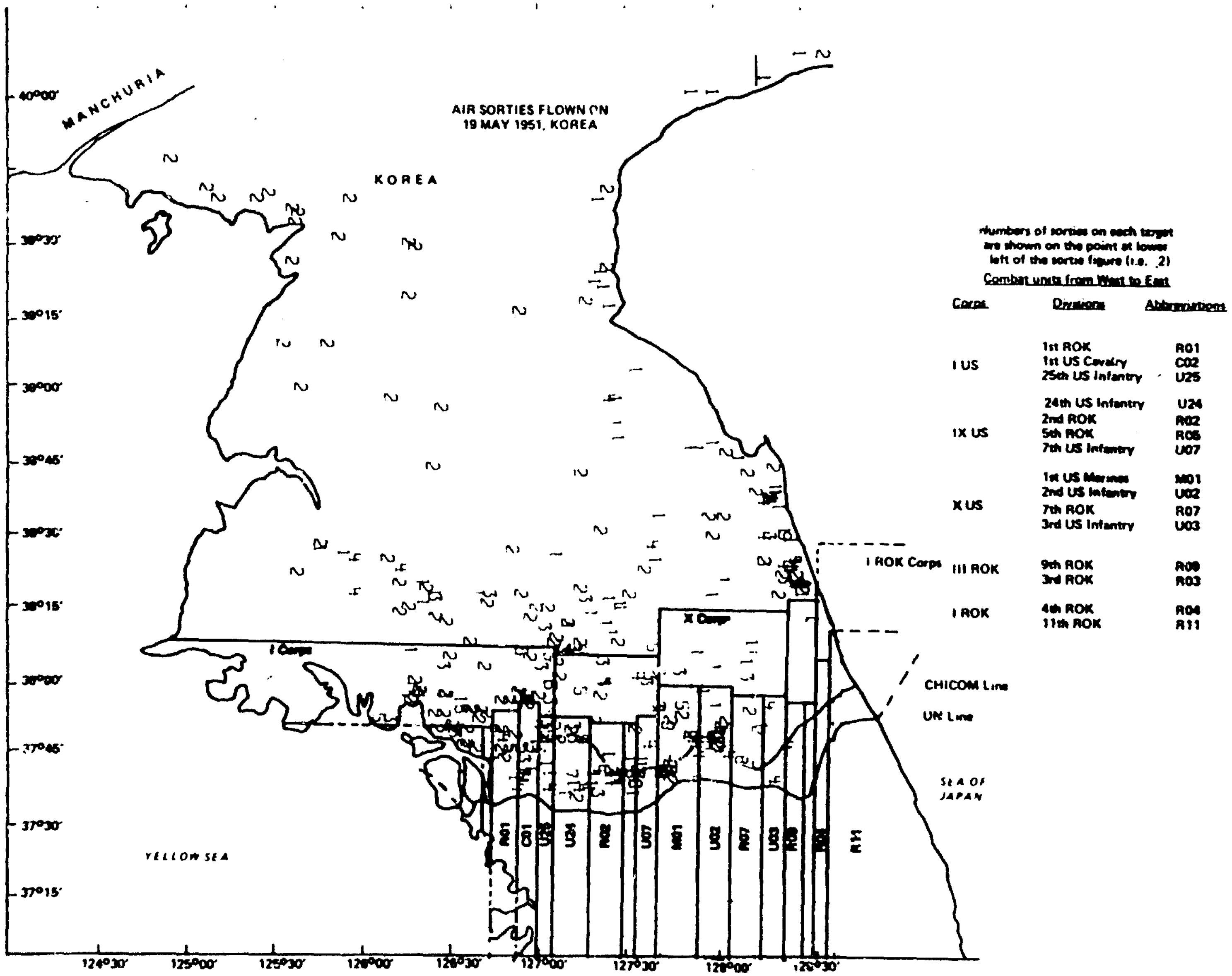

FIG 2: **SAMPLE OF AIR SORTIE DATA EXTRACTED FROM ORDNANCE IMPACT POINTS**

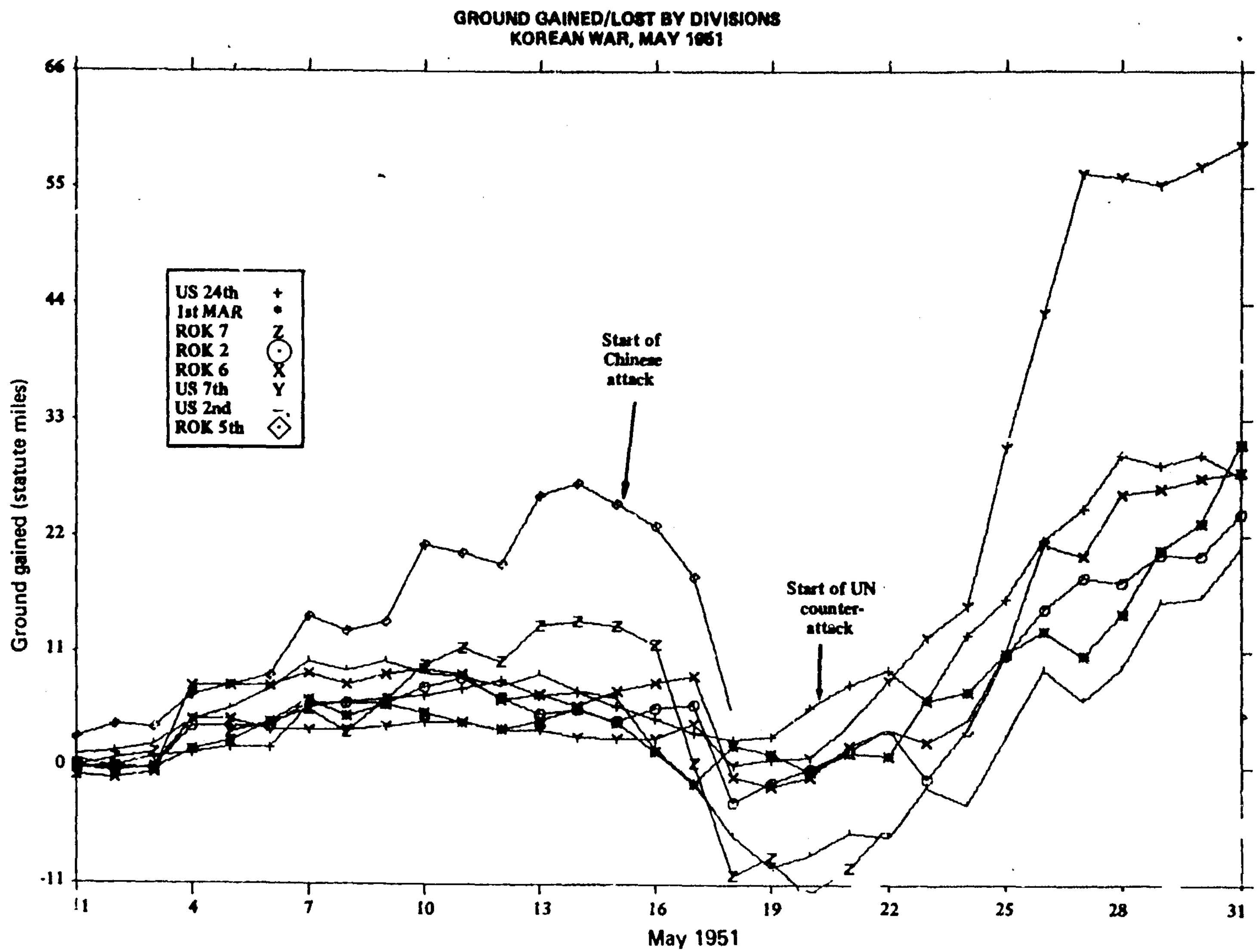

FIG 3: GROUND GAINED/LOST BY DIVISIONS, MAY 1951

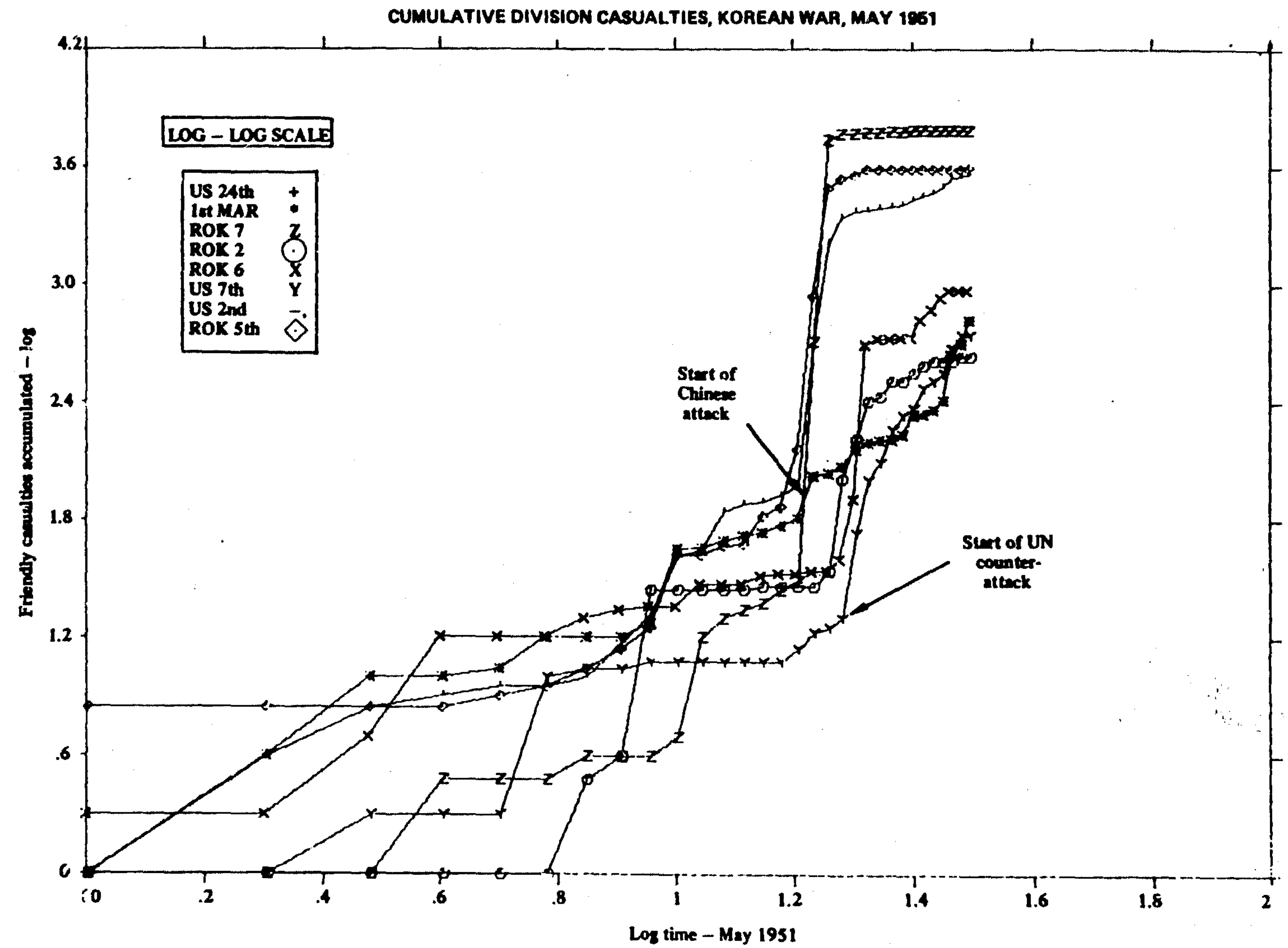

FIG 4: DATA FOR MAY 1951 – PLOTTED AS THE LOGARITHMS OF CASUALTIES AND TIME

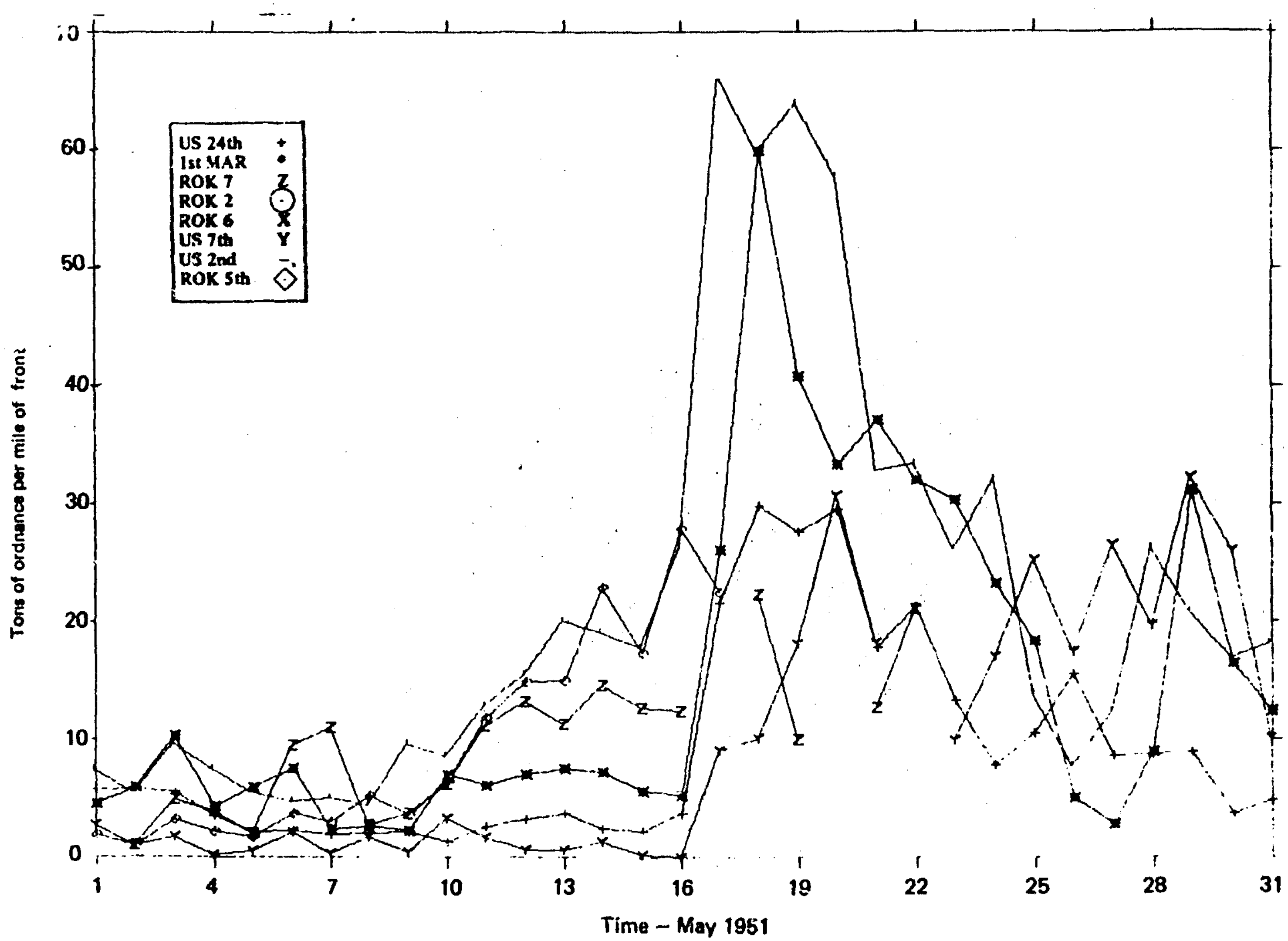

FIG 5: **TOTAL AIR AND ARTILLERY AMMUNITION TONS PER MILE FRONT KOREAN WAR, MAY 1951**

FIG: 6

REGRESSION COEFFICIENTS FROM DATA SORTED ON MISSION AND ON GROUND GAINED OR LOST (LANCHESTER LOG LAW I)

	Firepower equation	U. N. attack: Mission	U. N. attack: Ground gained	Enemy attack: Mission	Enemy attack: Ground gained	Patrol: Mission	Patrol: Ground gained by U.N.	Patrol: Ground gained by enemy
Enemy casualties	Separate	.39	.67	.61	.67	.81	.80	.83
	Total	.28	.61	.69	.75	.61	.62	.76
Friendly casualties	Separate	.07	.06	.08	.11	.22	.01	.06
	Total	.07	.06	.01	.11	.14	.01	.06

FIG: 7

REGRESSION EQUATIONS FOR CASUALTIES FOR TWO INTENSITIES OF CONFLICT WHEN GAINING GROUND
(See Series 41 in table E-2)

Conditions	Battle level	Equation	R^2
U. N. advancing	Intense	EC = -2295 + (47 AS_o + 750 HB_o)	.67 (.69)
Enemy advancing	Intense	EC = (.27 LA) EC = (.0004 total ord)	.68 (.74)
U. N. advancing	Static	EC = -58 + (.12 LA + 0.10 HA)	.78 (.80)
Enemy advancing	Static	EC = + (.10 LA + .16 HA)	.83 (.85)

* *Where ord. = tons of ordnance of the type expended.*

INDEX OF PROFESSIONAL PAPERS

NO.	SECTION	TITLE	AUTHOR(S)
1	INS	Static Models of Bank Credit Expansion	Brown/Lloyd
2	NAVWAG	The Sex-Differential in Canadian Unemployment Data	Lando
3	INS	A Dynamic Inventory Model with Delivery Lag and Repair	Brown/Corcoran/Lloyd
4	INS	A Moment Problem for Order Statistics	Kadane
5	INS	Optimal Whereabouts Search	Kadane
7	INS	The Continental Shelf Issue at the United Nations: Quantitative Content Analysis	Friedheim
8	INS	A Comparison of the Importance of Economics versus Non-Economic Factors Affecting the Residential Housing Market During the Two Decades Subsequent to World War II	Rose
9	INS	Existence of Excess Capacity at Naval Shipyards Prior to Escalation of Hostilities in Southeast Asia in 1964	Rose
11	INS	The Effect of Discrimination on Earnings: Evidence from Military Test Score Results	O'Neil
12	INS	Dynamic Models of Bank Credit Expansion Under Certainty	Brown/Lloyd
14	INS	Determination of the Optimal Investment in End Products and Repair Resources	Rose
15	INS	Computing the Expected End-product Service Time using Extreme Value Properties of Sampling Distributions	Rose
16	INS	A Study of Reparable Item Resupply Activities	Rose
17	INS	An Incremental Production Function for the End-item Repair Process	Rose
18	INS	Inventory and the Theory of the Firm	Rose
19	INS	A Decomposed Network Computation for End-product Repair Cost Curves	Rose
20	INS	Inventory Models with a Type of Dependent Demand and Forecasting with an Application to Repair	Brown/Corcoran/Lloyd

INDEX OF PROFESSIONAL PAPERS (Continued)

NO.	SECTION	TITLE	AUTHOR(S)
21	INS	Resource Allocation in a Sequential Flow Process	Silverman
22	INS	Israeli Reprisal Policy and the Limits of U.S. Influence	Gorlin
23	INS	An Aircraft Rework Cost-Benefit Model	Rose
24	INS	An Application of Network Analysis to the Determination of Minimum Cost Aircraft Pipeline Factors	Sutton/Lloyd
25	OEG	An Approach to Semi-Markov Processes	Saperstone
26	INS	The Reliability of a Complex System with Spares, Repair and Cannibalization	Brown/Corcoran
27	NAVWAG	Validation of Combat Models Against Historical Data	Feldman/Simon
28	INS	Quantitative Content Analysis of the United Nations Seabed Debates: Methodology and a Continental Shelf Care Study	Friedheim/Kadane
29	OEG	Controllability of Linear Oscillatory Systems using Positive Controls, I	Saperstone
30	INS	The Effect of Social Policy on the Social and Private Value of a Child	DeVany
31	INS	Time in the Budget of the Consumer	DeVany
32	INS	Fitting Korean War Data by Statistical Methods	Overholt
33	INS	A Theory of Household Demand and Labor Supply	DeVany

The ideas expressed in this paper are those of the authors. The paper does not necessarily represent the views of either the Center for Naval Analyses, the United States Navy or any other sponsoring agency. It has been reproduced by CNA as a courtesy to the authors, under the CNA Professional Development Program.

Center for Naval Analyses
an affiliate of the University of Rochester

1401 Wilson Boulevard Arlington, Virginia 22209 703/524-9400

20 May 1970

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

pfc 22 May 70

Gentlemen:

On 12 May 1970, copies of the following CNA Professional Papers were forwarded to you:

NC AD705349 CNA Professional Paper No. 32, "Fitting Korean War Data by Statistical Methods," by John L. Overholt, dated May 5, 1970

NC AD705350 CNA Professional Paper No. 33, "A Theory of Household Demand and Labor Supply," by Arthur DeVany, dated May 5, 1970

The attached statement was to have been printed inside the front cover. It is requested that this statement be placed inside the front cover. It has been prepared on self-adhesive labels which are easily inserted by removing the back of the label.

Sincerely,

Thomas J. Moore
Information & Document Center