

A Statistical Analysis of Historical Land Battles: What is Associated With Winning?

See theses of Yigit (2000), Coban (2001), Cakan (2003), and Gondal (2015)





Some Questions

- Who tends to have a higher casualty rate, attackers or defenders?
- How often do attackers win?
- What is the relationship between force ratio and victory?
- What factors are associated with winning land battles?
- Can we predict future combat outcomes?
- Do the answers to these questions change over time?

We are looking for relationships, patterns, and trends...

Context

"It is simply not wise to declare that history is not relevant and ignore the lessons of the past ... combat a peculiarly human pastime ... Only historical combat is real combat." – John R. Brinkerhoff on page xviii in the introduction to Numbers, Predictions, & War by Colonel T.N. Dupuy.

niorms





German Philosoph

Agenda

The CDB90G Data Set

• Summary Statistics

 Predicting Combat Outcomes with Classification and Regression Trees

Investigating the Questions With CAA's CDB90G Database

- 660 land battles from 1600 to 1982 within CDB90G
 - 552 battles were grouped into17 "logical conflicts" by Yigit (2000)
- 140+ attributes
 - Objective factors
 - Attacker/defender, force ratio, tank ratio, artillery ratio, primary defensive posture, etc.

Subjective factors

- Leadership, surprise, initiative, training, combat effectiveness, logistics, technology, etc.
- Weather and terrain factors



The 17 Groupings

- Thirty Years War (1620-1668)
- English Civil War (1642-1645)
- King Williams War (1689-1693)
- Austrian Succession War (1741-1745)
- Seven Year's War (1756-1760)
- American Revolutionary War (1775-1781)
- War of First Coalition (1792-1799)
- War of Second Coalition (1799-1800)
- Napoleonic Wars (1805-1815)
- U.S.-Mexico War (1846-1847)
- American Civil War (1861-1865)
- Franco-Prussian War (1870-1871)
- World War I (1914-1918)
- World War II (1939-1945)
- Arab-Israeli War (1948)
- Korean War (1950-1951)
- Arab-Israeli War (1973)

More CDB90G Details

RowNames	1600+ thru 1755	1755+ thru 1814	1814+ thru 1913	1913+ thru 1939	1939+ thru 1945	1945+ thru 2000	total
AUS	4	11	0	7	0	0	22
BR	1	12	7	24	26	0	70
CS	0	0	22	0	0	0	22
ENG	10	0	0	0	0	0	10
FR	14	28	5	12	1	0	60
GER	0	0	8	26	45	0	79
IS	0	0	0	0	0	49	49
00	29	20	26	17	3	23	118
PR	4	9	1	0	0	0	14
RUSS	0	0	3	4	0	0	7
SOV	0	0	0	3	24	0	27
USA	0	2	35	40	94	8	179
TOTAL	62	82	107	133	193	80	657

Table 1.Battles Per Period, Attacker.

RowNames	1600+ thru 1755	1755+ thru 1814	1814+ thru 1913	1913+ thru 1939	1939+ thru 1945	1945+ thru 2000	total
AUS	0	14	3	11	0	0	28
BR	2	10	4	7	11	0	34
CS	0	0	27	0	0	0	27
EG	0	0	1	0	0	26	27
FR	6	24	8	10	3	0	51
GER	0	0	1	69	110	0	180
IMP	12	0	0	0	0	0	12
IS	0	0	0	0	0	20	20
JAP	0	0	2	3	31	0	36
00	37	29	27	5	4	17	119
RUSS	1	4	5	10	0	0	20
SOV	0	0	0	5	6	0	11
SYR	0	0	0	0	0	14	14
TU	4	0	5	11	0	0	20
USA	0	1	24	2	28	3	58
TOTAL	62	82	107	133	193	80	657

Table 2. Battles Per Period, Defender.

More CDB90G Details

Table 9.Number of Battles in Different Centuries

NUMBER OF BATTLES IN CHRONOLOGICAL ORDER								
YEAR	Number of Battles	First Battle of Century	Last Battle of Century					
1600-1697	48	1600	1697					
1700-1799	65	1704	1799					
1800-1899	126	1800	1899					
1900-1982	425	1900	1982					
Total	664							



Figure 4. Battle Intensity over Time

Table 16.Summary Statistics Initial Force Ratio)
SUMMARY STATISTICS INITIAL FORCE RATIO AT	TACKER TO
DEFENDER	
Defender Wins Median	1.31
Defender Wins Mean	1.92
Defender Win Standard Deviation	2.23
Attacker Wins Median	1.66
Attacker Wins Mean	2.43
Attacker Win Standard Deviation	5.47



On the Data

- There are some missing data elements
- "Tie goes to the defender"
- For subjective factors, qualitative assessments are collapsed to:
 - Attacker Advantage (A)
 - Defender Advantage (D)
 - No Advantage (O)

Variation in Data

Characteristic	Туре	Year	Atkr & Dfdr	Value of	Ratio
Force Ratio Men	High :	1967	Egypt: Israel	17:1	57.1
(atkr:dfdr)	Low :	1945	Japan:USA	0.3:1	57.1
Force Ratio Artillery	High :	1945	USA:Japan	50:1	450.1
(atkr:dfdr)	Low :	1948	Israel:Syria	0.11:1	430.1
Mortar Density dfdr	High :	1943	Britain:Germany	132	720.1
(wpns/km)	Low :	1973	Egypt:Israel	0.19	/ 50:1
Artillery Density atkr	High :	1944	USA:Japan	444	2200.1
(wpns/km)	Low :	1948	Israel:Jordan	0.2	2200.1
Casualty Rate atkr	High :	1945	USA:Japan	96%	740.1
(% per day)	Low :	1944	Britain:Germany	0.13%	/40.1
Tank Loss Rate atkr	High :	1967	Israel:Syria	92%	150.1
(% per day)	Low :	1944	USA:Germany	0.63%	130.1
Advance Rate	High :	1967	Israel:Egypt	45	450.1
(km per day)	Low :	1945	USA:Japan	0.1	430.1

Let's Get Started: Who Has a Higher Casualty Rate?



Average Dispersion Over Time



Figure E.1. The Dispersions of the Campaigns. Note that the unit area is m².

How Often Does the Attacker Win?

Security against defeat implies defensive tactics; ability to defeat the enemy means taking the offensive — Sun Tzu

WINA	Unknown Outcome	Defender Wins	Draw	Attacker Wins
	(–9)	(–1)	(0)	(1)
Size	2	217	43	398

Attacker wins 60.4% — 64.7% if you don't include draws

Note: For model fits, we use 398 Attacker wins and 260 Defender wins

Force Ratio and Winning

The superiority in numbers is the most important factor in the result of a combat, only it must be sufficiently great to be a counterpoise to all the other co-operating circumstances. The direct result of this is, that the greatest possible number of troops should be brought into action at the decisive point — Clausewitz





More on Force Ratio (FR) and Winning

God is on the side of the big battalions — Napoleon

			FR = 3	3 or more	FR =	: 3 - 2.5	FR =	= 2.5 - 2	FR =	2 - 1.5	FR =	1.5 - 1.4	FR =	1.4 - 1.3	FR :	= 1.3-1	FR = ′	1 or less
	of Campaion	Total Number of	Number of	Number of battles attacker	Number	Number of battles attacker	Number of	Number of battles attacker										
	ane	battles	battles	wins	of battles	wins	of battles	wins	of battles	wins	of battles	wins	of battles	wins	of battles	wins	battles	wins
Year	740	N	n	a	n	а	n	а	n	a	n	а	n	а	n	a	n	a
1620	30 Years' War	18							1	1					4	4	13	12
1642	English Civil War	6					2	0	1	1							3	2
1689	King Williams' War	8							2	2			1	1	2	1	3	2
1741	Austrian-Success	7											1	1	1	0	5	4
1756	7 Years War	18	1	0	2	2			2	1					1	1	12	7
1775	American-Revol.	14	1	1					1	1					3	0	9	5
1792	War-of-1st-Coal.	14	3	2					2	2	1	1	2	1	2	2	4	1
1800	War-of-2nd-Coal.	7							2	1	1	1					4	1
1805	Napoleonic War	29			2	2	2	1	4	4	1	0	2	1	8	4	10	2
1846	US-Mexico	8			1	0									1	1	6	6
1861	American-Civil	49	3	3			6	3	12	3	3	2	1	1	10	2	14	2
1870	Farnco-Russian	10	2	2	1		1	1	3	3							3	3
1914	WWI	124	25	16	6	4	17	8	26	16	3	1	4	2	21	12	22	7
1944	WWII	191	66	47	25	17	30	23	27	17	4	1	10	4	13	5	16	5
1950	Korean war	11			1	0	2	2	2	2					1	1	5	4
1973	Arab-Israel 1973	29	5	1			3	1	3	0					3	1	15	11
	-																	
	Totals	543	106	72	38	25	63	39	88	54	13	6	21	11	70	34	144	74
	Average P(Attacker win	s given Fl	R)	0.6792		0.6579		0.6190		0.6136		0.4615		0.5238		0.4857		0.5139

Drawing the Picture





Close Air Support Sorties Ratio



Tank Ratio

"The best tank terrain is that without anti-tank weapons." — Russian Military Doctrine



Attacker Tactics/Defensive Posture

POST1 WINA	Delay (DL)	Fortified Defense (FD)	Hasty Defense (HD)	Prepared Defense (PD)	Withdrawal (WD)
-1	4	72	129	55	0
1	16	107	164	107	2
TOTALS	20	179	293	162	2

Table 1.The Defender's Primary Posture

PRIA1 WINA	Double Envelopment (DE)	Defensive Offensive Plan (DO)	Single Envelopment (EE)	Frontal Attack (FF)	River Crossing (RC)
- 1	5	0	11	238	5
1	14	1	30	324	22
Size	19	1	41	562	27

Table 1.The Attacker's Primary Tactics

Who Wins (Tactics and Posture)?

Posture Tactics	Delay	Fortified Defense	Hasty Defense	Prepared Defense	Withdrawal
Double Envelopment	0/0	1/4 0.25	8/10 0.80	5/5 1.00	0/0
Offensive Defensive Plan	0/0	0/0	1/1 1.00	0/0	0/0
Single Envelopment	1/1 1.00	5/7 0.71	17/24 0.71	6/8 0.75	1/1 1.00
Frontal Attack	10/13 0.77	96/160 0.60	136/256 0.53	80/131 0.61	1/1 1.00
River Crossing	1/2 0.50	5/7 0.71	2/2 1.00	13/15 0.87	0/0

Table 1.The Attacker's Chances of Victory Given the Attacker's Tactics
and the Defender's Posture

Moving to Subjective Factors: Surprise

Strike the enemy at a time or place or in a manner for which he is unprepared — FM 100-5 Operations

SURPA WINA	Α	D	0
-1	32	14	206
1	124	0	249
TOTALS	156	14	455

Table 1.The Distribution of the SurpriseVariable, "SURPA"

Relative Air Superiority Advantage

Control of air gives commanders the freedom to conduct successful attacks that can neutralize or destroy an enemy's war fighting potential — FM 100-5 Operations

AEROA WINA	Α	D	0
-1	67	38	147
1	162	19	192
TOTALS	229	57	339

Table 1.The Distribution of the Relative Air Superiority
Variable, "AEROA"

Relative Leadership Advantage

The mere presence of Napoleon on a battlefield was worth 40,000 men — Wellington or Blucher (according to Dupuy)

LEADA WINA	А	D	Ο
-1	13	88	151
1	155	12	206
TOTALS	168	100	357

Table 1.The Distribution of the Relative LeadershipAdvantage Variable, "LEADA"

Relative Combat Effectiveness Advantage

It's the unconquerable soul of man, and not the nature of the weapon he uses, that ensures victory — General George S. Patton

CEA WINA	Α	D	Ο
-1	22	55	175
1	120	27	226
TOTALS	142	82	401

Relative Combat Effectiveness Advantage (CEA)

Relative Training Advantage

We train like we fight and we fight like we train — U.S. Army

TRNGA WINA	Α	D	Ο
-1	24	50	178
1	87	52	234
TOTALS	111	102	412

Table 1.The Distribution of the Relative Training
Advantage Variable, "TRNGA"

Relative Morale Advantage

Morale is to the Physical as three is to one — Napoleon Morale is the single greatest factor in successful wars — Eisenhower

MORALA WINA	Α	D	Ο
-1	22	8	222
1	110	2	261
TOTALS	132	10	483

Table 1.The Distribution of the Relative MoraleAdvantage Variable, "MORALA"

Relative Logistics Advantage

Logistics cannot win a war, but its absence or inadequacy can cause defeat — FM 100-5 Operations



Table 1.The Distribution of the Relative LogisticsAdvantage Variable, "LOGSA"

Relative Momentum Advantage

The energy developed by good fighting men is as the momentum of a round stone rolled down a mountain thousands of feet in height — Sun Tzu



Table 1.The Distribution of the Relative Momentum
Advantage, "MOMNTA"

Relative Intelligence Advantage

If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle. — Sun Tzu

INTELA WINA	Α	D	0
-1	7	39	206
1	73	8	292
TOTALS	80	47	498

Table 1.The Distribution of the Relative IntelligenceAdvantage Variable, "INTELA"

Relative Initiative Advantage

Initiative: The power or ability to begin or to follow through energetically with a plan or task; enterprise and determination — Webster



Table 1.The Distribution of the Relative Initiative
Advantage Variable, "INITA"

Terrain Effect

terra1.1 WINA	Flat (F)	Rugged (G)	Rolling (R)	Other (O)
- 1	44	48	163	3
1	69	83	235	3
TOTALS	113	131	398	6

Table 1.The First Terrain Descriptor

Correlations Between Variables

	WINA	fR	tank	arty	fly	CEA	LEADA	MORALA	INTELA	TECHA	TRNGA	LOGSA	SURPA	AEROA	INITA	MOMNTA
WINA	1.00	0.13	0.11	0.10	0.12	0.32	0.52	0.26	0.33	0.07	0.17	0.21	0.27	0.21	0.45	0.22
fR	0.13	1.00	0.18	0.53	0.41	- 0.13	- 0.14	0.22	0.06	0.00	- 0.13	0.08	0.04	0.17	- 0.02	0.11
tank	0.11	0.18	1.00	0.21	0.50	0.04	- 0.08	0.14	- 0.03	0.33	0.02	- 0.07	- 0.05	0.14	- 0.01	- 0.15
arty	0.10	0.53	0.21	1.00	0.31	0.02	- 0.10	0.17	0.04	0.09	0.07	0.05	0.01	0.12	0.00	0.05
fly	0.12	0.41	0.50	0.31	1.00	0.03	- 0.04	0.06	- 0.07	0.21	0.05	0.06	- 0.13	0.28	- 0.02	- 0.09
CEA	0.32	- 0.13	0.04	0.02	0.03	1.00	0.54	0.10	0.06	0.18	0.68	0.02	0.13	0.09	0.28	0.14
LEADA	0.52	- 0.14	- 0.08	- 0.10	- 0.04	0.54	1.00	0.06	0.23	0.06	0.39	0.08	0.21	0.11	0.40	0.11
MORALA	0.26	0.22	0.14	0.17	0.06	0.10	0.06	1.00	0.09	0.15	- 0.08	0.18	0.06	0.28	0.20	0.26
INTELA	0.33	0.06	- 0.03	0.04	- 0.07	0.06	0.23	0.09	1.00	0.06	0.06	0.19	0.41	- 0.04	0.29	0.10
ТЕСНА	0.07	0.00	0.33	0.09	0.21	0.18	0.06	0.15	0.06	1.00	0.14	0.10	0.04	0.18	0.12	0.12
TRNGA	0.17	- 0.13	0.02	0.07	0.05	0.68	0.39	- 0.08	0.06	0.14	1.00	0.09	0.09	0.02	0.15	0.21
LOGSA	0.21	0.08	- 0.07	0.05	0.06	0.02	0.08	0.18	0.19	0.10	0.09	1.00	0.06	0.18	0.10	0.24
SURPA	0.27	0.04	- 0.05	0.01	- 0.13	0.13	0.21	0.06	0.41	0.04	0.09	0.06	1.00	- 0.04	0.28	- 0.03
AEROA	0.21	0.17	0.14	0.12	0.28	0.09	0.11	0.28	- 0.04	0.18	0.02	0.18	- 0.04	1.00	0.14	0.15
INITA	0.45	- 0.02	- 0.01	0.00	- 0.02	0.28	0.40	0.20	0.29	0.12	0.15	0.10	0.28	0.14	1.00	0.27
MOMNTA	0.22	0.11	- 0.15	0.05	- 0.09	0.14	0.11	0.26	0.10	0.12	0.21	0.24	- 0.03	0.15	0.27	1.00

Forecasting Battle Winners

In your deliberations, when seeking to determine the military conditions, let them be made the basis of a comparison, in this wise:

- (1) Which of the two sovereigns is imbued with the Moral law?
- (2) Which of the two generals has most ability?
- (3) With whom lie the advantages derived from Heaven and Earth?
- (4) On which side is discipline most rigorously enforced?
- (5) Which army is stronger?
- (6) On which side are officers and men more highly trained?
- (7) In which army is there the greater constancy both in reward and punishment?
- By means of these seven considerations I can forecast victory or defeat.

– SunTzu,

We will use Classification and Regression Trees

Why Classification and Regression Trees?

- Easy to understand
- Straightforwardly handle a set of predictors that contains a mix of numeric variables and qualitative factors.
- Nonparametric
- Tree-based models are adept at capturing nonadditive behavior (e.g., interactions naturally emerge).
- Tree-based models are invariant to monotone re-expressions of predictor variables, so the precise form in which these appear in a model formula is irrelevant.

Three Models Built

Model 1 — Objective factors: Force ratio, "fR", CAS sorties ratio, "fly", tank ratio, "tank", artillery ratio, "arty", cavalry ratio, "cav", attacker's primary tactical scheme, "PRIA1", defender's primary defensive posture, "POST1"

Model 2 — Model 1 + Relative factors: Surprise, "SURPA", air superiority in the theater, "AEROA", combat effectiveness, "CEA", leadership advantage, "LEADA", training advantage, "TRNGA", morale advantage, "MORALA", logistics advantage, "LOGSA", momentum advantage, "MOMNTA", intelligence advantage, "INTELA", technology advantage, "TECHA", initiative advantage, "INITA"

Model 3 — Model 2 + terrain and weather factors: 9 total factors

Sub-setting the Data

SUBSET NO	SUBSET	SIZE	TRAINING SET	SIZE	TEST SET	SIZE
1	1600 -1847	164	1600 - 1799	109	1799 - 1847	55
2	1805 - 1918	260	1805 - 1915	178	1916 - 1918	82
3	1920 - 1945	202	1920 - 1944	131	1944 - 1945	71
4	1940 - 1982	223	1940 - 1948	150	1950 - 1982	73
5	1600 - 1982	658	1600 - 1944	435	1940 - 1982	223

Validating the Models

Base Model = Attacker Wins (60 percent correct)

	Test1	Test2	Test3	Test4	Test5
All outcomes	55	82	71	73	223
Clear-cut outcomes	45	60	32	50	126

Table 1.The Sizes of Test Sets, and Test Sets with
Only Clear-Cut Outcomes

Objective Factors on Subset 5 (Model 1: the whole data

-1



Figure 36. Model 1.6 for Subset 5

Summary of Goodness of "Objective" Models

Subset	Misclassification Rate of the Training Set	Misclassification Rate of the Test Set	Misclassification Rate of the Test Set with Clear-cut Outcomes		
Subset 1 Yrs. 1600-1847	0.30	0.55	0.60		
Subset 2 Yrs. 1805-1918	0.31	0.57	0.58		
Subset 3 Yrs. 1920-1945	0.31	0.34	0.13		
Subset 4 Yrs. 1940-1982	0.21	0.29	0.36		
Subset 5 Yrs. 1600-1982	0.32	0.42	0.37		
Data set Yrs. 1600-1982	0.38	NA	NA		

Misclassification Rates of the Models for Objective Variables

Model 2 Including All Factors



Figure 37. Model 2.1 for Entire Data Set

Adding Subjective Factors Improves Fits Dramatically

	Misclassification Rate of the Training Set	Misclassification Rate of the Test Set	Misclassification Rate of the Test Set with Clear-Cut Outcomes
Subset 1 Yrs. 1600-1847	0.15	0.20	0.16
Subset 2 Yrs. 1805-1918	0.12	0.24	0.20
Subset 3 Yrs. 1920-1945	0.17	0.30	0.13
Subset 4 Yrs. 1940-1982	0.29	0.34	0.26
Subset 5 Yrs. 1600-1982	0.15	0.26	0.15
Data Set Yrs. 1600-1982	0.18	NA	NA

Important Variables and Misclassification Rates Over Time

- Approach
 - Use 125 battles to predict the next one
 - Found by trial and error to work well
 - Results in 533 predictions
 - Average rates consist of 30 predictions
 - To assess trends, a sliding window of five battles is used

Size	50	75	100	125	150	175	200
Misclassification rate	0.32	0.29	0.27	0.27	0.28	0.28	0.38

Table 34.Misclassification Rates versus Training Set Size

Clear Pattern in Misclassification Rates





Importance of Variables





Accounting For All Splits (Weighted)





Is There a Nationality Affect?



Figure 1. Tree Model for the Battles Before WWI.

World War I Battles



Figure 13. Model 2 Battles of World War I.

World War II Battles



Figure 14. Model 3, Battles of World War II.

Focus on Four Countries

	1600+	1913+	1939+	1945+	
	thru 1913	thru 1939	thru 1945	thru 2000	total
USA	37	40	94	8	179
BR	30	24	26	0	80
GER	22	26	45	0	93
IS	0	0	0	49	49

Table 3.Battles Per Period, Attacker.

	1600+	1913+	1939+	1945+	
	thru 1913	thru 1939	thru 1945	thru 2000	total
USA	25	2	28	3	58
BR	16	7	11	0	34
GER	1	69	110	0	180
IS	0	0	0	20	20

Table 4.Battles Per Period, Defender.

Proportion of Battles Won by Attacker



Figure 2. Proportion of Battles Won By Attacker.

Force Ratios of Attacking Countries



Figure 3. Force Ratios of Attacking Countries.

Table 5.Force Ratio Averages.

Summary Thoughts...

- Retrospectively, classification and regression trees do okay (sometimes nearly 85%) at predicting combat outcomes
- Classification models fit only to objective factors do not predict well
- The importance of leadership
 Especially with hindsight
- Recent emergence of objective and technical factors

Questions or Comments?

- Gondal, Bilal S., Major, Pakistan Army, "Statistical Analysis of Warfare: Identification of Winning Factors with a Focus on Irregular Warfare," M.S. in Operations Research, September 2015.
- Cakan, Ali, First Lieutenant, Turkish Army, "Determining the Importance of Nationality on the Outcome of Battles Using the Classification Trees," M.S. in Operations Research, June 2003.
- Coban, Muzaffer, First Lieutenant, Turkish Army, "Predicting Battle Outcomes With Classification Trees," M.S. in Operations Research, December 2001.
- Yigit, Faruk, First Lieutenant, Turkish Army, "An Exploratory Analysis of Historical Land Battles," M.S. in Operations Research, December 2000.



