National Defence University
Series 1: Research Publications No. 44

Modelling of Military Capabilities, Combat Outcomes and Networked Systems with Probabilistic Methods

Vesa Kuikka

Vesa Kuikka

Modelling of military capabilities, combat outcomes and networked systems with probabilistic methods

To be presented, with the permission of the Research Council of National Defence University, for public criticism for the degree of Doctor of Military Sciences in auditorium Itälinnake at the National Defence University, Santahamina, Helsinki, on August 20th 2021 at 12 o'clock.

NATIONAL DEFENCE UNIVERSITY
HELSINKI 2021

MAANPUOLUSTUSKORKEAKOULU
JULKAISUSARJA 1: TUTKIMUKSIA NRO 44

NATIONAL DEFENCE UNIVERSITY
SERIES 1: RESEARCH PUBLICATIONS NO. 44

MODELLING OF MILITARY CAPABILITIES, COMBAT OUTCOMES AND NETWORKED SYSTEMS WITH PROBABILISTIC METHODS

VESA KUIKKA

NATIONAL DEFENCE UNIVERSITY
HELSINKI 2021

Vesa Kuikka: *Modelling of military capabilities, combat outcomes and networked systems with probabilistic methods*
Maanpuolustuskorkeakoulu
Julkaisusarja 1: Tutkimuksia nro 44
Väitöskirja /

National Defence University
Series 1: Research Publications No. 44
Doctoral dissertation

Author:	Licentiate of Philosophy Vesa Kuikka
Supervisor:	Professor Jouko Vankka National Defence University
Pre-examiners:	Professor Kai Virtanen Aalto University
	Professor Eero Hyry Tampere University
Opponents:	Professor Eero Hyry Tampere University
	PhD Ilja Honkonen Finnish Meteorological Institute

Recent publications in PDF format: *http://www.doria.fi/handle/10024/73990*

ISBN 978-951-25-3194-3 (Pbk.)
ISBN 978-951-25-3195-0 (PDF)
ISSN 2342-9992 (print)
ISSN 2343-0001 (web)

Maanpuolustuskorkeakoulu
National Defence University

PunaMusta Oy
Tampere Finland 2021

To my two aunts and uncle

ABSTRACT

This study is about the mathematical modelling of three important concepts. In selecting these three areas, we had certain general principles in mind, in particular: they should cover a wide range of concepts in the military domain and they should be central in the areas of military planning and operations. As a consequence of these guidelines, the three military concepts had to be made abstract to cover a wide range of military applications. Selecting from the set of highly conceptual concepts, that are central and important, is not unambiguous but it is easy to come up with good candidates. We selected the following three areas as our modelling objectives: military capabilities, combat outcomes and networked defence systems. These three concepts satisfy both the broadness and centrality criteria.

Our aim to cover a wide range of applications has also another dimension: using a common mathematical methodology enables us to develop quantitative models in different conceptual operational areas. The advantage of this approach is to have the same basis in the entire work we are going to perform.

Because our goal is to have a common mathematical methodology in describing military concepts, we end up with a natural choice of probabilistic modelling. We don't have proof that probability theory is the only possibility in our case; however, it is probably the only one that meets the requirement of quantitative modelling. As a consequence, this approach allows comparing the modelling results in the selected application areas. Further, it also enables extending the models to more detailed structures or other application areas in the military domain. This shows that probability theory can be applied to a wide range of modelling problems in military studies.

All the proposed models have been tested with numerical examples. These examples show that the models provide reasonable and consistent results. Probabilistic models are demonstrated in the three case studies of military capabilities, combat outcomes and networked systems. The model proposed for modelling combat outcomes is also used to model combat effectiveness during the engagement in three combats where the time-dependent empirical data is available.

Keywords: military capability, combat modelling, combat equation, combat duration, networked defence system, optimal sensor placement, optimal number of system units, probabilistic modelling, operations research

TIIVISTELMÄ

Työn otsikkona on ”Sotilaallisten suorituskykyjen, taisteluiden lopputulosten ja verkottuneiden järjestelmien mallintaminen todennäköisyyslaskennan menetelmillä”. Tämä tutkimus käsittelee kolmen tärkeän käsitteen matemaattista mallintamista. Näitä kolmea aihealuetta valitessa noudatimme erityisesti joitakin yleisiä periaatteita: aiheiden tulee kattaa laaja käsitejoukko sotilaallisella alalla ja niiden tulee olla keskeisiä sotilaallisen suunnittelun ja operatiivisen toiminnan kannalta. Näiden periaatteiden seurauksena valittavien sotilaallisten käsitteiden oli oltava abstrakteja, jotta ne kattaisivat laajan joukon sotilaallisia sovelluksia. Korkean käsitteellisen tason keskeisten ja tärkeiden aiheiden valinta ei ole yksikäsitteistä, mutta on helppo löytää hyviä ehdokkaita sellaisiksi. Valitsimme mallintamisen kohteiksi seuraavat kolme osa-aluetta: sotilaalliset suorituskyvyt, taisteluiden lopputulokset ja verkottuneet puolustusjärjestelmät. Nämä kolme aihetta toteuttavat sekä laajuuden että keskeisyyden vaatimukset.

Tavoitteella kattaa laaja sovellusten kenttä on myös toinen ulottuvuus: yhteisten matemaattisten menetelmien käyttäminen mahdollistaa kvantitatiivisten mallien kehittämisen erilaisille toiminnallisille alueille. Tämän lähestymistavan etuna saavutetaan koko tutkimuksen perusta on yhteinen.

Koska tavoitteenamme on käyttää yhteistä matemaattista metodologiaa sotilaallisten käsitteiden kuvaamiseen, päädymme luonnollisena valintana todennäköisyysmalleihin. Emme todista, että todennäköisyyslaskenta on tässä tapauksessa ainoa vaihtoehto, mutta se on todennäköisesti ainoa, joka toteuttaa vaatimuksen kvantitatiivisesta mallintamisesta. Tämä tekee mahdolliseksi mallintamisen tulosten vertailun valituilla aihealueilla. Edelleen se mahdollistaa mallien laajentamisen yksityiskohtaisempiin rakenteisiin tai muille sotilaallisille soveltamisalueille. Todennäköisyyslaskentaa voidaan soveltaa laajaan joukkoon mallintamisongelmia sotilaallisessa tutkimuksessa.

Kaikki ehdotetut mallit on testattu numeeristen esimerkkien avulla. Nämä esimerkit osoittavat, että malleista saatavat tulokset ovat järkeviä ja johdonmukaisia. Todennäköisyysmallit todennetaan kaikissa kolmessa tapauksessa: sotilaalliset suorituskyvyt, taisteluiden lopputulokset ja verkottuneet järjestelmät. Mallia joka kuvaa taisteluiden lopputuloksia käytetään myös mallintamaan taisteluiden aikaista tehokkuutta kolmessa taistelussa, joista on saatavilla havaintotietoa ajan funktiona.

Avaisanat: sotilaallinen suorituskyky, taistelumalli, taistelun kaava, taistelun kesto, verkottuneet järjestelmät, optimaalinen sensoreiden sijoittaminen, optimaalinen järjestelmäyksiköiden lukumäärä, todennäköisyysmalli, operaatiotutkimus

ACKNOWLEDGEMENTS

I would first like to thank my supervisor professor Jouko Vankka at the National Defence University of Finland for the guidance during the writing of this research work.

I would like to acknowledge my colleagues Marko Suojanen and Juha-Pekka Nikkarila from the Finnish Defence Research Agency for their contributions to the research articles used as sources in this study.

I am also grateful to pre-examiners Prof. Eero Hyry and Prof. Kai Virtanen for their valuable comments on an earlier version of the manuscript.

I would like to recognize the assistance that I received from the study guidance staff at the National Defence University of Finland who offered me help during the process of completing my studies in military technology.

Vesa Kuikka

Helsinki 27.4.2021

LIST OF APPENDED JOURNAL ARTICLES

This manuscript consists of an introduction and seven journal articles, which have been published in peer-reviewed journals, referred to as (P I) – (P VII) in the text.

P I Modelling the Impact of Technologies and Systems on Military Capabilities
Vesa Kuikka and Marko Suojanen
Journal of Battlefield Technology, 17(2), 2014, pp. 9-16

P II Dependency of Military Capabilities on Technological Development
Vesa Kuikka, Juha-Pekka Nikkarila and Marko Suojanen
Journal of Military Studies, 6(2), 2015, pp. 29-58

P III Number of System Units Optimizing the Capability Requirements through Multiple System Capabilities
Vesa Kuikka
Journal of Applied Operational Research, 8(1), 2016, pp. 26-41

P IV A Combat Equation Derived from Stochastic Modeling of Attrition Data
Vesa Kuikka
Military Operations Research, 20(3), 2015, pp. 49-69

P V Decision Boundaries Used to Model Probability of Victory and Duration of Combats
Vesa Kuikka
Journal of Applied Operational Research, 9(1), 2017, pp. 67-81

P VI Metrics for Networked Systems Design in a Network-Centric Warfare Context
Vesa Kuikka
Journal of Battlefield Technology 10(3), 2007, pp. 27-36

P VII Optimal Sensor Placement in the Network Structure from a Defence Point of View
Vesa Kuikka and Juha-Pekka Nikkarila
Journal of Information Warfare, 19(2), 2020, pp. 46-61

The author's contributions to the respective appended journal articles have been:

P I Modelling and most of the research design and most of the writing. (MS has participated in writing the article and in designing the questionnaire concerning the development of capabilities. MS has defined the three scenarios used in the questionnaire.)

P II Modelling, research design and writing. (J-PN and MS have commented on the article manuscript. MS has designed the questionnaire concerning the development of technology areas.)

P III The sole work of the author

P IV The sole work of the author

P V The sole work of the author

P VI The sole work of the author

P VII Modelling and research design and most of the writing. (J-PN has commented on the article manuscript and written section 'Background: Changes in Cyberspace and the Closing of National Infrastructure Networks'.)

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

C_j	Task capability of task j
$A_{i,j}$	Development of technology area i on task capability j
$D_{i,k}$	Interdependence of technology area i on technology area k
Δ_k	The parameter describing faster or slower technological development of technology area k with respect to the original evaluation in the model (e.g. $\Delta_k = 0$ no development, $\Delta_k = 1$ original evaluation, $\Delta_k = 2$ 100% faster development)
M	Necessary system of systems in an initial ensemble of a system of systems (see Fig. 2.2)
K	An alternative system of systems in the initial ensemble of a system of systems (see Fig. 2.2)
S_i	Alternative systems added to the initial ensemble of a system of systems (see Fig. 2.2)
X_m	System capability of system M (see Fig. 2.2)
X_k	System capability of system K (see Fig. 2.2)
X_i	System capability of the system S_i (see Fig. 2.2)
p_0	The capability of the initial ensemble of a system of systems $p_0 = X_m X_k$
p_1	Increment to the initial capability value p_0 of systems M and K on the capability area level after adding system S_1 (see Fig. 2.2)
p_{12}	Increment to the initial capability value p_0 of systems M and K on the capability area level after adding system S_1 and S_2 (see Fig. 2.2)
$p_{1\ldots i}$	Increment to the capability value of the system of systems $M, K,$ $S_j, j = 1, \ldots, i-1$ on the capability area level after adding system i (see Fig. 2.2)
$A_{1\ldots i}$	System capability of the system of systems $S_j, j = 1, \ldots, i$ (see Fig. 2.2)
n_i	Number of system units of systems i
n_i'	The threshold for system i, $n_i \geq n_i'$ saturated state and $n_i < n_i'$ non-saturated state
f	Functionality $f = 1, \ldots, F$
p_f	The capability of functionality f in the initial ensemble of a system of systems $p_f = X_{mf} X_{kf}$
M_f	Necessary functionality in an initial ensemble of a system of systems (see Fig. 2.2)
K_f	Alternative functionality in an initial ensemble of a system of systems (see Fig. 2.2)
$A_{1\ldots \mathcal{I} f}$	The capability of functionality f of systems $i = 1, \ldots, \mathcal{I}$, $A_f \equiv A_f(n_1, \ldots, n_{\mathcal{I}}) \equiv A_{1\ldots \mathcal{I} f}$
X_{if}	Capability value of functionality f provided by all system i
x_{if}	Capability value of functionality f provided by one system unit of system i
$\mathcal{L}$	Lagrangian function

λ	Lagrangian multiplier
C	Cost of a system of systems
$g_i(n_i)$	Cost of the number of n_i system units of system i (e.g. $g_i^{(lin)}$ and $g_i^{(dis)}$)
h_i	Cost of one system unit
δ_i	Tuning parameter for discount amount of system i in $g_i^{(dis)}(n_i) = h_i n_i^{1-\delta_i}$
T	Time at the end of the battle $t = T$
x_0, x_u	Attacker force size at time $t = 0$ and $t = T$
y_0, y_d	Defender force size at time $t = 0$ and $t = T$
W_t	Standard Brownian motion with mean zero and variance t
B_t	Brownian motion with drift $\hat{\mu}$ and variance σ^2, $dB_t = \hat{\mu}dt + \sigma dW_t$
X_t	Geometric Brownian motion describing the attacker attrition process
Y_t	Geometric Brownian motion describing the defender attrition process
σ^2	The variance of geometric Brownian motion, volatility σ
σ_A	The volatility of the attacker attrition process
σ_D	The volatility of the defender attrition process
μ	The drift of geometric Brownian motion
μ_A	The drift of the attacker attrition process including the Itô term, $\mu_A = \ln(x_u/x_0)$
μ_D	The drift of the defender attrition process including the Itô term, $\mu_D = \ln(y_d/y_0)$
$\hat{\mu}_A$	The drift of the attacker attrition process
$\hat{\mu}_D$	The drift of the attacker defender process
M_A	The slope of the attacker decision boundary
M_D	The slope of the defender decision boundary
α	Parameter for tuning the attacker decision boundary, $M_A = -(1-\alpha)\mu_D - \alpha\mu_A$
β	Parameter for tuning the defender decision boundary, $M_D = -(1-\beta)\mu_A - \beta\mu_D$
S	Ratio defined by $S = (x_u y_0)/(y_d x_0)$
P_A	Probability for the attacker to win the battle, $P_A = \left((S^{\alpha-1})^{-\lvert m_A\rvert} - A^{-\lvert m_A\rvert}\right)/\left(1 - A^{-\lvert m_A\rvert}\right)$
P_D	Probability for the defender to lose the battle, $P_D = \left(\left(S^{1-\beta}\right)^{-\lvert m_D\rvert} - D^{-\lvert m_D\rvert}\right)/\left(1 - D^{-\lvert m_D\rvert}\right)$
m_A	$m_A = 1 - 2(\hat{\mu}_A + M_A)/\sigma_A^2 = -2(\mu_A + M_A)/\sigma_A^2$
m_D	$m_D = 1 - 2(\hat{\mu}_D + M_D)/\sigma_D^2 = -2(\mu_D + M_D)/\sigma_D^2$
A	$A = S^{\alpha-1}(1 + \mathfrak{p}_A) - \mathfrak{p}_A$

D	$D = S^{1-\beta}(1 + \mathfrak{p}_D) - \mathfrak{p}_D$
$\mathfrak{p}_A$	The parameter determines the initial position of the attacker lower decision boundary
$\mathfrak{p}_D$	The parameter determines the initial position of the defender upper decision boundary
$E(T)$	Expected duration of a battle
d_{xu}	The parameter in the definition of the upper attacker decision boundary in Model 2
d_{xd}	The parameter in the definition of the lower attacker decision boundary in Model 2
d_{yu}	The parameter in the definition of the upper defender decision boundary in Model 2
d_{yd}	The parameter in the definition of the lower attacker decision boundary in Model 2
$\mathbb{M}$	Stochastic matrix (or Markov matrix), matrix elements $M_{i,j}$ describe the transition from i to j
n	Number of nodes in the system
p_z	Probability of failure
p_t	Probability of replacement
p_r	Probability of repair
π	Probability of a functioning system
π_0	Probability of a non-functioning system, $\pi + \pi_0 = 1$
π_i	Limiting probability of i functioning nodes
$Q(n)$	The ratio of uptime to downtime of the system of systems with n systems
p	Probability of a functioning element (node or link)
p_{s-t}	Probability of functioning connection between nodes s and t (Eqs. (69) – (72))
$\mathcal{G}$	Network structure (topology) of nodes and links
$r(\mathcal{G})$	Reliability of network $\mathcal{G}$, the average probability of connectivity between pairs of nodes
N_L	Number of links in network $\mathcal{G}$
N_s	Number of services in network $\mathcal{G}$
N	Number of nodes in the network
$\mathbb{P}$	Matrix $\mathbb{P}$, Matrix elements $p_{s,t}$ describe the probability of functional connection from node s to node t ($p_{s,t}$ corresponds to p_{s-t} in Eqs. 69-72)
$\mathbb{S}_z$	Matrix $\mathbb{S}_z$ of service z
$\mathbb{C}$	Matrix $\mathbb{C}$ describing the usefulness of connections and services, matrix elements $C_{s,t}, s, t = 1, \dots N$
$\mathbb{U}$	Matrix $\mathbb{U}$, overall utility, $\mathbb{U} = \mathbb{C} \circ \mathbb{P}$, Matrix elements $U_{s,t}, s, t = 1, \dots, N$

$C_{s,t}$	The values of matrix elements are obtained by aggregating the link weights across each service
$C(s,t)$	Information spreading probability from node s to node t (see (Kuikka 2018a))
$\mathcal{C}$	Sum of $C(s,t)$ over $s,t = 1,\dots,N$
B_b	Same as $C(s,t)$ but node n is removed from the network
b_n	Betweenness measure defined as $b_n = (\mathcal{C} - B_n)/\mathcal{C}$
$P_{i,L(i,\mathfrak{c})}$	Intermediate step i in calculating the element $C(s,t)$
$\mathfrak{c}$	Index for the two combined paths $\mathfrak{c} = 1,2$ and their mutual path $\mathfrak{c} = 3$
$L(i,\mathfrak{c})$	Path length in iteration i for index $\mathfrak{c}$
W_L	The product of all link and node weights along the path
$M_c(\pi_c)$	The objective function for sensor placement using a centrality measure
π_c	Penalty parameter in the objective function $M_c(\pi_c)$
$M_b(\pi_b)$	The objective function for sensor placement using a betweenness measure
π_b	Penalty parameter in the objective function $M_b(\pi_b)$

CENTRAL CONCEPTS

Algebraic connectivity	Algebraic connectivity (also known as Fiedler value or Fiedler eigenvalue) of a graph G is the second-smallest eigenvalue (counting multiple eigenvalues separately) of the Laplacian matrix of G.
Analytic hierarchy process	The Analytic Hierarchy Process (AHP) is a structured technique for organising and analysing complex decisions, based on mathematics and psychology.
Attrition	Gradually making something weaker and destroying it, especially the strength or confidence of an enemy by repeatedly attacking it.
Capability area	Collections of like military activities functionally grouped to support capability analysis, strategy development, investment decision making, capability portfolio management, and capabilities-based force development and operational planning.
Capability (systems engineering), System capability	A Capability, in the systems engineering sense, is defined as the ability to execute a specified course of action. A capability may or may not be accompanied by an intention. The term is used in the defence industry but also in private industry.
Connectivity	The ability of a computer, program, device or system to connect with one or more others.
Delphi method	The Delphi method is a forecasting process framework based on the results of multiple rounds of questionnaires sent to a panel of experts.
Equation	In mathematics, an equation is a statement that asserts the equality of two expressions, which are connected by the equals sign "=".
Markov matrix	In mathematics, a Markov matrix is a square matrix used to describe the transitions of a Markov chain. Each of its entries is a nonnegative real number representing a probability. It is also called a probability matrix, transition matrix, substitution matrix or stochastic matrix.
Mathematical optimisation	Mathematical optimisation is the selection of the best element (with some criterion) from some set of available alternatives.
Model	Something such as an object, plan, or set of rules used to show what something else is like or how it works.

Networked computers, Networked systems	Networked computers are connected so that they can share information.
Network criticality	Network criticality is a graph-theoretic metric that quantifies network robustness, and that was originally designed to capture the effect of environmental changes in core networks.
Network topology, Network structure	Network topology is the topological structure of a network and may be depicted physically or logically. It is an application of graph theory wherein communicating devices are modelled as nodes and the connections between the devices are modelled as links between the nodes.
Optimal stopping	In mathematics, the theory of optimal stopping or early stopping is concerned with the problem of choosing a time to take a particular action, to maximise an expected reward or minimise an expected cost.
Protection	1) DoD: Preservation of the effectiveness and survivability of mission-related military and non-military personnel, equipment, facilities, information, and infrastructure deployed or located within or outside the boundaries of a given operational area. 2) JCA: The ability to prevent/mitigate adverse effects of attacks on personnel (combatant/non-combatant) and physical assets of the United States, allies and friends. It was decomposed into Prevent, Mitigate, and Research and Development.
Resilience	The ability to be happy, successful, etc. again after something difficult or bad has happened.
Situational awareness	Situational awareness or situation awareness is the perception of environmental elements and events concerning time or space, the comprehension of their meaning, and the projection of their future status.
Stochastic	A stochastic process or system is connected with a random probability.
Stopping rule	A termination decision rule, Stopping boundary.
System of systems	A system of systems is a collection of task-oriented or dedicated systems that pool their resources and capabilities together to create a new, more complex system that offers more functionality and performance than simply the sum of the

	constituent systems.
Technology	The use of scientific knowledge or processes in business, industry, manufacturing, etc.
Utility	Ability to satisfy a particular need, usefulness.

LIST OF ABBREVIATIONS

AHP	Analytic Hierarchy Process
APC	Armoured Personnel Carriers
CBEA	Capabilities-Based Engineering Analysis
DoD	U.S. Department of Defense
ERP	Enterprise Resource Planning (ERP) is the integrated management of main business processes, often in real-time and mediated by software and technology.
IEEE	Institute of Electrical and Electronics Engineers
INFORMS	Institute for Operations Research and the Management Sciences
JCA	Joint Capability Areas (CAs) are collections of DoD capabilities functionally grouped to support capability analysis, strategy development, investment decision making, capability portfolio management, and capabilities-based force development and operational planning. JCAs are logically broken down from higher-level capability categories to further scope, bound, and clarify capability categories by providing greater granularity to facilitate detailed analysis or allow better mapping of resources to capabilities. The number of tiers/levels required to decompose a JCA down to its component capabilities is not a constant across the JCAs. JCAs identify the major functional areas of near and far-term challenges.
JETCD	Joint Experimentation, Transformation and Concepts Division (U.S.)
MCDA	Multiple-Criteria Decision Analysis
MORS	Military Operations Research Society
NATO	North Atlantic Treaty Organisation
OR	Operations Research
QJM	Quantified Judgment Model
SDN	Software-Defined Network
SEM	Structural Equation Modeling
UAV	Unmanned Aerial Vehicle

TABLE OF CONTENTS

1.

INTRODUCTION

Our goal is to find a limited set of quantitative methods and models that can describe high-level functional concepts in the military domain. The objective of quantitative research is to develop and employ mathematical models, theories and hypotheses about phenomena (Booß-Bavnbek 2003). To this end, we develop probabilistic models in the areas of studying military capabilities, combat outcomes and operability of networked systems. These representative topics are all related to each other and a model describing one of them can have common characteristics and dependencies with others. Many new possibilities and opportunities can be identified in models considering two or three of these areas of military studies. We propose novel mathematical models and corresponding formulas in all these areas (Fig 1.1). The models are extendable to include more systems or cover more detailed system structures. These methods can provide new insights to understanding interrelationships between different aspects of combats, capabilities and networks. In the very end, the purpose of this study is to provide useful tools that can be helpful in military processes like planning military capabilities or in justifying decision-making.

In developing probabilistic models, we still have alternatives. As we are developing mathematical models for high-level military concepts, using analytic expressions is a natural choice. More specifically, the modelling results of this study can be expressed in closed-form mathematical formulas. Simulation methods are often used to compute numerical results of probabilistic models. Numerical simulations are needed if the model is not expressed in a closed or analytic form or if the analytic expressions of the model are so complicated that numerical simulations are easier to perform. Often closed-form expressions are more compact and expressive than simulation algorithms. This is important because we aim to understand the mathematical expressions to have a tractable chain of deduction from the military concepts to the modelling results.

Mathematical formulas can be used in specific scenarios with numerical input values for the explanatory (or independent) variables. Response (or dependent) variables predict the outcome of the modelled quantities or they explain the variation of the explanatory variables. The analyst using the model aims to determine the values of controllable explanatory variables to optimise a measure of effectiveness. Uncontrollable variables cannot be changed. In a modelling project, visualisations of response variable values in both analytic and simulation models can use similar methods. Using the output data of the model can be adequate for an end-user. However, limitations and application areas of the model should be understood, and in this respect, it may be helpful to go through the mathematical derivation of the model.

1.1. Modelling Military Capabilities, Combats and Networked Systems

The outline of this study is the following. First, we discuss general concepts of operations research, purposes of modelling and probability-based modelling. Next, we have a short overview and summary of related work on the areas of capability modelling, combat modelling and modelling networked systems. Later, in Sections 2, 3 and 4, we present research articles written by the author. Lastly, the main results in Section 5 and conclusions in Section 6 end the introductory part. The original research articles upon which this study is built are collected at the end of this study.

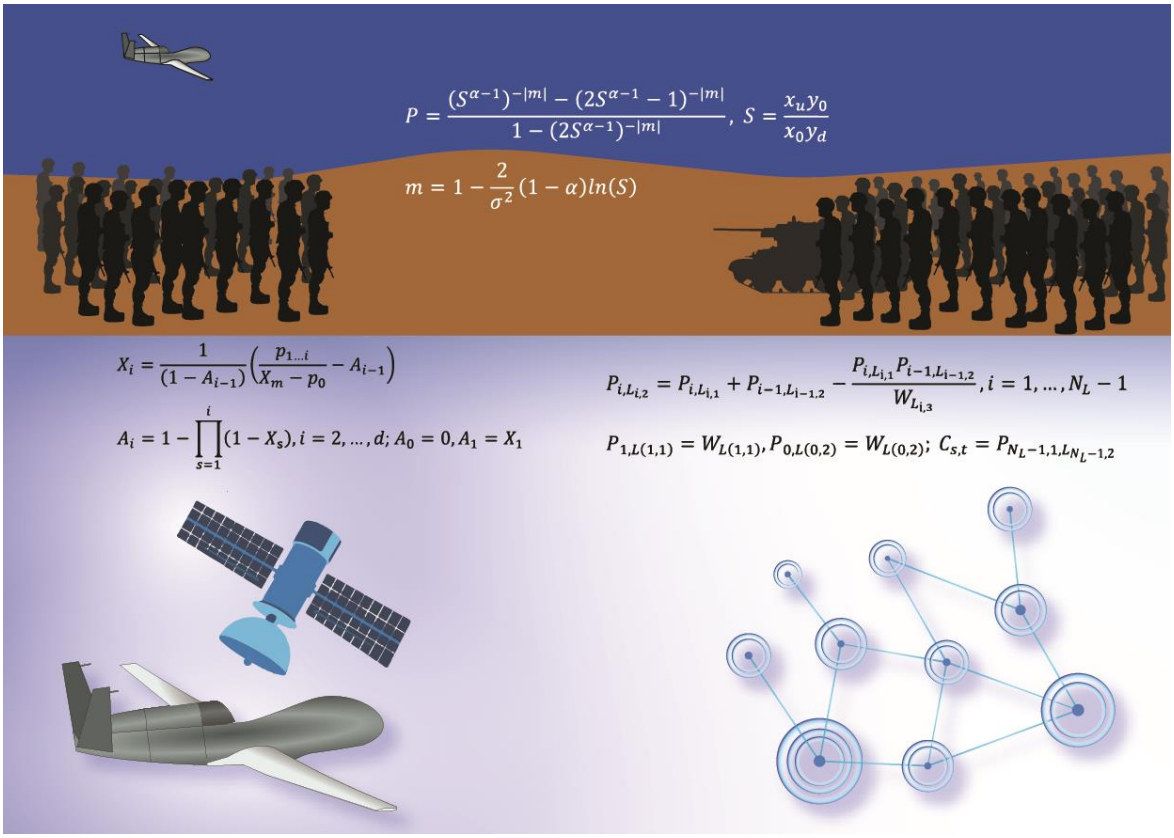

Fig. 1.1. Schematic illustration of three areas of modelling: Capabilities (left), Combats (top) and Networks (right).

1.1.1. Operations Research

Operations research, also called operational research, relies on the application of scientific methods in the management and administration of organised military, governmental, commercial, and industrial processes[1] (Carter et al. 2019). The central objective of operations research is optimisation: to perform the best possible way under the given circumstances. Operations research is a discipline that applies analytical methods to support better decision making. Further, the term operational analysis is used in the military as an intrinsic part of capability development, management and

[1] Definition of 'operations research' in (Morse & Kimball 1951): Operations research is a scientific method of providing execute departments with a quantitative basis for decisions regarding the operations under their control.

assurance (Scala & Howard II 2020; National Research Council 2006). In particular, operational analysis forms a part of the combined operational effectiveness and investment appraisals, which help make better decisions about defence acquisitions. (Morse & Kimball 1951; Washburn & Kress 2009; Bracken et al. 1995)

1.1.2. Why Model?

A question that sometimes arises: Why model? Joshua Epstein (Epstein 2008) points in his essay that the mathematical model at least gives a hypothesis that can be analysed and falsified, in contrast to a purely verbal argument that may be too general or too equivocal to be tractable. He also notes that a central purpose of simple mathematical models is to illuminate 'core dynamics', the understanding of which is a necessary beginning and complement to more elaborate modelling or simulations. A strong argument for the importance of simple models is that they necessarily capture the central truths hidden within more complex models, which refine rather than supersede the simpler models. On the other hand, modelling is a method to clear up misconceptions and ideas to ascertain that the author and reader will always see things from the same point of view[2]. The motivation of many combat models in literature has been the observation that similar core dynamics pervade various approaches to describe attrition in warfare. (Epstein 2008; Edmonds et al. 2019)

Niall MacKay's lecture (MacKay 2015) elaborates on some enduring misconceptions about modelling. One of these is that the goal is always prediction. The lecture distinguishes between explanation and prediction as modelling goals and offers sixteen reasons other than prediction to build a model. It also challenges the common assumption that scientific theories arise from and 'summarise' data, when often, theories precede and guide data collection; without theory, in other words, it is not clear what data to collect. Among other things, it also argues that the modelling enterprise enforces habits of mind essential to freedom. (MacKay 2015; Edmonds et al. 2019)

1.1.3. Probability-Based Modelling

A probabilistic method or model is based on the theory of probability or the fact that randomness plays a role in explaining events. The opposite is deterministic, which tells us something that can be described or predicted exactly. Probabilistic models use random variables and probability distributions in describing events or phenomenon.

[2] (Clausewitz 1832) Carl von Clausewitz wrote: "But the first business of every theory is to clear up conceptions and ideas which have been jumbled together, and, we may say, entangled and confused; and only when a right understanding is established, as to names and conceptions, can we hope to progress with clearness and facility, and be certain that author and reader will always see things from the same point of view. Tactics and strategy are two activities mutually permeating each other in time and space, at the same time essentially different activities, the inner laws and mutual relations of which cannot be intelligible at all to the mind until a clear conception of the nature of each activity is established. He to whom all this is nothing, must either repudiate all theoretical consideration, or his understanding has not as yet been pained by the confused and perplexing ideas resting on no fixed point of view, leading to no satisfactory result, sometimes dull, sometimes fantastic, sometimes floating in vague generalities, which we are often obliged to hear and read on the conduct of War, owing to the spirit of scientific investigation having hitherto been little directed to these subjects."

While a deterministic model gives a single possible outcome for an event, a probabilistic model gives a probability distribution as a solution. In this study, for example, the concept of military capability will be defined with the probability of success of the intended objective. (Florescu 2015)

One example of probabilistic methods is the use of relevance trees. Relevance trees have been widely used as a technology forecasting method. The principle behind using the relevance tree is to evaluate systematically all the related technologies leading to the success of the intended objective. The relevance tree presents information in a hierarchical structure. A hierarchy begins with objectives which are further decomposed into activities and further into tasks. As one descends in the hierarchy, the details increase at every level. The entries describe the preceding level completely when taken together at each higher level. Also, all the activities and tasks depicted should be mutually exclusive. (Coates & Glenn 2009)

1.2. Overview of the Modelling Areas of This Study

Short overviews of the three modelling areas of military capabilities, combat outcomes and networked systems are provided in this section. Sections 2, 3 and 4 provide a more detailed discussion about the original articles (P I – P VII). Related work of the modelling areas is summarised in Section 1.3.

1.2.1. Overview: Modelling Military Capabilities

In this section, we will present a short overview of the concept of military capabilities. Also, we bring up some methods in the literature that are related to our work. They can supplement our methods, or in some cases, they can be used as alternative approaches.

The concept of capability commonly is used in the planning and modelling of military power. Military capability is a composition of different existing and potential strengths. The concept of capability has been used to express the level of will, amount of troops and armament. Combat capability is a generalised characteristic of quantity and quality of forces and assets. Measuring military capability[3], and the difficulty of quantifying military capability in a single definite measure, have been discussed in (GAO 1985; Craig et al. 1991; Biltgen & Mavris 2006; Moore et al. 1991; Saperstein 1994; Jürgen 2003).

Traditionally, capability as a general concept consists of material, manpower, logistic and leadership viewpoints. It is a major issue to combine these aspects to understand the strength of military forces against different adversaries and changing environmental situations. Often military capability is defined only from the material viewpoint

[3] https://www.rand.org/content/dam/rand/pubs/monograph_reports/MR1110/MR1110.ch7.pdf, Accessed 21.1.2021

(Dupuy 1985) considering only one of the capability areas. It is straightforward to compare the number of aeroplanes, tanks and ships. However, this approach has obvious drawbacks: Just counting these equipment does not take into consideration the technological level of implementations and their utility in potential combat situations. Secondly, it is possible to use different alternative equipment to achieve comparable levels of capability, for example, by using tanks or aeroplanes.

The concept of capability in the military context has been standardised in many countries. In the US the Joint Capability Areas (JCA) is a set of nine standardised definitions that cover the complete range of military activities (JETCD 2009). JCA consists of a hierarchical structure of functionalities under main capability areas such as situational awareness, protection and engagement. The Finnish Defence Forces have adapted a similar capability structure together with the general concept of capability as achieving a predefined level of effects in a specified situation (Kosola 2013). Commonly used definitions of capability are closely related to the probabilistic definition of our study. One definition is: "Capability is the ability to achieve a specified wartime objective like win a war or battle, or destroy a target." The current U.S. Department of Defence definition (DoD 2020)[4] and NATO definition (NATO 2019)[5] are quite similar. "The ability to create an effect through the employment of an integrated set of aspects categorized as doctrine, organisation, training, materiel, leadership development, personnel, facilities and interoperability."

Military capability is measured by its effectiveness in accomplishing the objectives assigned. We require the definitions of capability areas and their sub-capabilities not to overlap each other and to cover the whole range of capabilities. We can use any specification of capabilities fulfilling these two requirements. Typically definitions include a hierarchical structure of sub-capabilities with increasing details of functionalities (Kosola 2013). The requirements for capability specifications are logically inherited from all levels of the hierarchy. Many of the capability definitions obey the requirements of no overlapping and full coverage of capabilities. This is analogous to the traditional method of normalisation in conceptual analysis. A further consequence follows from probability theory: capabilities, or probabilities in our definition, can be assumed statistically independent due to the no-overlapping property (P I).

---

[4] (DoD 2020) Capability: The ability to achieve a specified wartime objective (win a war or battle, destroy a target set). It includes four major components: force structure, modernization, readiness, and sustainability.

a) Force structure: Numbers, size, and composition of the units that comprise US defense forces; e.g., divisions, ships, air wings.
b) Modernization: Technical sophistication of forces, units, weapon systems, and equipment.
c) Unit readiness: The ability to provide capabilities required by the combatant commanders to execute their assigned missions. This is derived from the ability of each unit to deliver the outputs for which it was designed.
d) Sustainability: The ability to maintain the necessary level and duration of operational activity to achieve military objectives. Sustainability is a function of providing for and maintaining those levels of ready forces, materiel, and consumables necessary to support military effort.

[5] (NATO 2019) The ability to create an effect through employment of an integrated set of aspects categorized as doctrine, organisation, training, materiel, leadership development, personnel, facilities, and interoperability.

The general concept of military capability is an abstract construct that is not directly applicable to mathematical calculations. In this study, we define the concept of capability as the probability of a successful mission or operation. The corresponding system capability is defined as the probability of a successful system operation. In our definition, non-material aspects, such as tactics and will on opposing sides, affect capabilities in a complex manner. Definitions of capabilities can be taken as mathematical interpretations of the general concept of capability as reaching a predefined level of effects in a particular scenario (Amer et al. 2013). One problem in capability modelling is how to determine relationships between individual system capabilities and higher-level capability areas. In literature, heuristic models have been proposed where relations are described by weighting factors. These models may have difficulties in consistently defining terms and concepts because different quantities or alternative systems are combined in mathematical formulas. (P I)

1.2.2. Overview: Combat Modelling

Understanding relationships between the outcome of a battle and factors like force sizes of opposing sides is a fundamental question in military studies[6]. For an army leader to be able to explain combat effects or predict combat outcomes more accurately would be vital. Any improvements in tools and theories or other methods of increasing general awareness would be most welcome. Historically force size is the most important factor but force quality, tactics, casualties, duration, surprise, advance rates and victory are also important (Hartley 2001). In modern warfare, technological systems, not only weaponry but also communication and information systems have increasing importance (Hartley 2020). The roles of information and networking have increased substantially in the last few decades by UAVs and other electromechanical systems. Changes are likely to continue, especially in asymmetrical conflicts. Tactics have also changed from historical battles: the distributed and networked operation mode has been adopted by many armed forces. (Bracken et al. 1995; Low 1995)

Many historical battle data samples have been aggregated and used (Hartley, 2001). The aggregate database covers a wide range of force ratios, and while emphasising 20th Century battles, has reasonable coverage back to 1600. The database has numerical values for force sizes at the beginning of a battle and casualties at the end of a battle for both sides of conflicts. Also, the empirical data includes durations for each battle. The database of almost 600 lines indicates the attacker and the winner for each battle.

Methods used to study combat effects are numerous and cover many fields of military sciences, information analysis, mathematics, psychology and social sciences. Much of

[6] "The most crucial problem affecting the analytical or simulation modeling of combat at virtually any level of intensity is identified as the lack of a body of knowledge, or a theory, that defines the relationship between combat modeling and the conduct of war in the real world." Lawrence J. Low, "Theater-Level Gaming and Analysis Workshop for Force Planning," September 1977.

the work has been heuristic in the sense that results are not presented in quantitative terms. Quantitative methods include the use of mathematical formulas or simulation methods. One line in the long history of studying combats is trying to write down an analytic combat equation, preferably a simple one, describing combat effects, for example predicting the winner of a battle. However, only a few closed-form formulas exist today. (Bracken et al. 1995; Low 1995)

One may argue that no simple formula can describe complex relationships in warfare, and modern warfare is becoming even more complex. In our view, this does not exclude the need to understand the big picture of the complex system of combats. If there are common features in all combats or specific types of combats, these features should be recognized and studied with appropriate methods. Secondly, mathematical models and results from the analysis may give a benchmark for comparing and characterising combats. This would give a basis for more detailed and comprehensive models. Identifying the most important factors and understanding what are the interrelationships between them is the basic work necessary before a more detailed theory can be developed. Comparing predictions of models and empirical observations provide insight into the validity of assumptions of the models. (Epstein 2008; Edmonds et al. 2019)

Quantitative mathematical formulas should be understandable and derived and expressed with real-world variables. In this respect, some combat models are phenomenological and some are based on quantifiable variables. Examples for the latter are the Lanchester equations (Lanchester 1914; Osipov 1915). F. Lanchester and M. Osipov concurrently and independently devised a series of differential equations to demonstrate the power relationships between opposing forces. M. Osipov published a five-part series of articles in 1915 in the Russian journal Voenniy Sbornik. Lanchester equations are expressed in terms of force sizes and attrition rates.

Examples of phenomenological formulations are the Helmbold relationship (Helmbold 1987) and the Willard equation (Willard 1962). They use force sizes as explanatory variables but also incorporate some phenomenological parameters without a real word interpretation. These parameters are typically fitted with empirical data. Models having both types of variables, phenomenological and empirical, are also common.

Both types of methods are useful, but our goal in this study is to give a quantitative interpretation or explanation to all expressions, or at least, provide guidelines for future studies about how to construct a more detailed quantitative model. Decision boundary, or decision rule, in (P V) is an example that can be considered as a phenomenological concept, but in theory, it could be defined exactly, for example, based on enquiries among combat leaders. There is a real word meaning behind the concept of the decision boundary, but the problem is how to define it exactly, and how to relate it with other relevant quantities of the problem.

1.2.3. Overview: Modelling Networked Systems

Network Science is a growing field of research (Barabási A.-L. 2016, Liu et al. 2020). One definition of network science is the study of network representations of physical, biological and social phenomena leading to predictive models of these phenomena. In addition to specialised methods of network modelling, networks can be studied with general methods of computer science and mathematics. Graph theory alone is a broad theoretical branch of mathematics. Much of the current research of applied network science draws on statistics, data analysis, data mining and many other fields of mathematics. These methodologies can be used together with methods based on network structure or independently, especially, when the structure and properties of the investigated system are not fully known. (Newman 2010; Newman 2012; Barabási 2016)

Applications cover areas such as telecommunication, computer networks, transportation, biological networks, cognitive and semantic networks and social networks. Generally speaking, any system having dependencies or relationships between its sub-systems or elements can be modelled with a network structure. One more narrow characterisation of network science is an academic field that studies complex networks considering distinct elements represented by nodes (or vertices) and the connections between the elements as links (or edges). Dynamics on networks has been traditionally modelled with simulations and random walks. Analytical formulas have been developed for specific network structures such as scale-free networks (Newman 2010; Barabási 2016). Complex topological structures make analytical solutions difficult or in many cases intractable.

Networks are modelled with nodes and links between nodes. Links describe connections, interactions or other relations between nodes. Links can be bi-directional having different strengths. Strengths can describe different things such as the strength of social ties or capacity of communication lines. Most real-world networks have a non-symmetric topology. In social networks, for example, all members of the network are not connected and the strengths of connections are different.

Nodes in a network can have different roles as central influencers, mediators or peripheral nodes. The degree is the simplest closeness centrality measure. The degree centrality of a node is defined as the number of nodes directly connected to it. This is a local measure and it does not take into account the position of the node in the network structure. The closeness centrality of a node measures how central or influential the node is to other nodes in the network. Betweenness centrality measures the role of a node as a proxy between other nodes in the network. Closeness and betweenness centrality measures have many variants and their usability depend on the research question of a particular application. Mathematical models provide more centrality measures, for example, the eigenvector centrality (Martin et al. 2014) and the

influence centrality measure (Kuikka 2018a; Kuikka 2021). In-centrality and out-centrality can be defined in directed networks. Normally, the betweenness centrality of a node is not defined for inward and outward directions. Review article (Gómez 2019) provides eight definitions and descriptions of the most commonly used centrality measures. Several other centrality measures and their variants exist in the literature. (Newman 2010; Gómez 2019)

1.3. Related Work

In this section, we review related work and concepts in the literature. Capability modelling, combat modelling and networked systems are discussed in the next sub-sections.

Several mathematical models exist in the area of capability modelling but their approach is different from the one proposed in this section. We discuss some general modelling approaches commonly used in the literature. Application of mathematical modelling to military capability areas is not found in the research literature, except the articles of the author of this study. Related work in the areas of technology forecasting and technological development are discussed because in our work military capabilities are investigated in these applications.

References concerning combat modelling are from a long period but recently the subject has not been researched much. Especially, only a few analytic predictive or prescriptive combat equations have been presented in the literature and they are based on deterministic modelling of combats (Lanchester 1914; Osipov 2015; Kress 2020; Washburn & Kress 2009; Helmbold 1987; Willard 1962). The approach in this study is to derive an analytic combat equation based on the theory of stochastic analysis in mathematics.

Network resilience has been studied extensively in military and non-military contexts. Recently a review article has been published (Liu et al. 2020) discussing resilience function and regime shift of complex systems in different domains, such as ecology, biology, social systems and infrastructure. The two specific topics in this study consider the use of Markov matrix formalism and optimal sensor placement in a network topology. The Markov matrix is a standard tool for describing system transitions (Karlin & Taylor 1975) whereas there does not exist a substantial amount of research regarding optimal sensor placement.

1.3.1. Related Work: Capability Modelling

The Quantified Judgment Model, Trade-off Analysis and Military Utility are three concepts related to capability modelling. As technology forecasting and technological development are used as application areas for capability modelling, these topics are

discussed in this section. More literature references of technology forecasting and technological development have been presented in (P II).

Quantified Judgment Model.
The Quantified Judgment Model (QJM) is an example dealing with many of the issues of comparing quantitative and qualitative aspects of military forces. QJM is based on the numbers of troops and weapons. Comparing the weighted sum of the two opposing sides predicts the winner of a battle or war. The model takes into account, in addition to weaponry and force sizes, human factors such as leadership, troop morale and operative factors like terrain and weather. Parameters of the model have been calculated based on theoretical and empirical knowledge. The power potential value for both sides is calculated and the ratio of the two values predicts the victor. (Dupuy 1985)

Trade-off Analysis.
The trade-off analysis technique is a closely related method in the decision process of choosing among the variety of alternatives in a situation of uncertainty. These decisions can range from changing your manners as a leader, selecting personnel for tasks and choosing the optimal systems for an operation. The trade-off analysis process has the following steps: understand the context, define the alternatives, define the criteria, set the criteria weights, set the scores, analyse the results and take the decision. These phases should be carried out before the modelling effort, to some extent, depending on the functional and technical requirements for the actual research question. Detailed information on the trade-off analysis methods and their phases can be found in (Parnell 2017).

General methodologies like Multiple-Criteria Decision Analysis (MCDA) (Tzeng & Shen 2017), Structural Equation Modeling (SEM) (Cheung 2015), Analytic Hierarchy Process (AHP) (Saaty 1990) will be discussed later in the context of future research in Section 7.

Military Utility.
The concept of military utility has been proposed for the study of technological systems in military operations (Andersson et al. 2015). The concept helps support communication and effective decision-making within the defence community. Military utility deals with selecting military technology and how to use that technology. It affects performance on the battlefield and the sustainment of capabilities over time.

The military utility is a measure of military effectiveness, suitability and affordability in a specific context. The concept is derived through conceptual analysis (Goertz & Mahoney 2012) and is based on related concepts of social sciences and system engineering. The descriptive capacity of the concept is useful for military decision-making processes regarding technology forecasting, defence planning, development and deployment of defence systems. (Andersson et al. 2015)

Military effectiveness is a measure of the overall ability to accomplish a mission when technology is used by representative personnel in an environment where operational use of military force is planned or expected. Military suitability is a measure of how well a technological system or concept, within a specified context, can produce military performance with consideration of other external factors. Affordability is a measure of compliance to the number of resources a military actor has allocated in a time frame defined by the context. (Andersson et al. 2015)

By using the concept of military utility and a system approach it is possible to explain how military capabilities are constituted and affected by developments in technology, by different use of technology and how military command levels are affected differently. Many factors are influencing the assessment and apart from providing a common methodology, it brings forth primary factors and relationships which must be considered in the assessment.

Technology Forecasting

There are four elements in a technology forecast: a time horizon, a specific technology, some parameters to the technology and a probability statement about the outcome (Fleisher & Bensoussan 2007). Forecasting methods suitable for a particular situation are determined by the available data and the objective of the forecast. The key point is to characterise the forecasted quantity or quantities as the most suitable methods may be different according to the properties of the quantity.

Usually, forecasting results are given as a function of time. Forecasted results may be economic, weather or demographic quantities, for example. A common procedure is to calculate the output quantities based on empirical or historical input data. Future forecasted values are forecasted by using historical values of input variables. One may use several mathematical methods: factor analysis, principal component analysis and linear regression just to mention a few. (Walk 2012)

We aim to forecast capability changes and reveal the origins of capability enhancements arising through the deployment of new systems and technologies and by the development of technology in existing systems. Structured modelling of systems and the probability-based description of capabilities can be a realistic forecasting method when compared with the usual linear regression or exponential extrapolation methods. We mention two other methods that have been used successfully in the literature, a nonparametric method called Technology Forecasting with Data Envelopment Analysis (TFDEA) (Inman et al. 2006) and linear regression methods and compare these methods with the structural forecasting method of this study.

One aspect in comparing forecasting methods is how reliable and detailed input data is used in modelling. Too many input variables and too detailed models may result in biased results over a longer time horizon. Detailed models may be appropriate for short term forecasting. Essential in long term forecasting is to discover the development trends. TFDEA and linear regression methods can be compared from these two

angles. The TFDEA method is not modelling the details of the system, however, the correct choice of input variables is essential in both methods. When forecasting extreme events, the choice of the input variable is particularly important.

The introduction time of a new version of a fighter is an example of a situation where the forecasted variable is a measure of an extreme event (Inman et al. 2006). The introduction date of a new fighter model depends on the best, or extreme, development achievements of construction, stealth, motor and control system technologies. TFDEA is a successful method for predicting such kinds of extreme events. TFDEA uses technology frontier analysis and optimal combination of existing technologies for forecasting, where the optimal combination is calculated by the linear programming method. Linear regression has also been used for the same forecasting problem (Martino 1993). Linear regression is an approach for modelling the relationship between a scalar dependent variable and one or more explanatory variables. TFDEA method had a better agreement with the observed dates of deployment (Martino 1993). Unlike TFDEA, linear regression as a method is not optimised for extrapolation of extreme events or taking into account the optimised combination of technological development.

Technological Development

Optimal resource allocation (Bridgeman 2013) is vital in many fields of business and public planning. Ideas presented in this paper are demonstrated in a military context (Neaga et al. 2009; Kerr et al. 2006; Smith & Oosthuizen 2011; Smith et al. 2012; Bordon et al. 2014) but the same principles can be applied in many other fields (Webb 2008). Resource reservation models in public health emergencies have been studied in (Bridgeman 2013). The main research question is how to allocate the limited available budgeted funds to alternative systems and equipment. The optimal configuration of systems is determined by the maximum capability (Kudryavtsev et al. 2014; Kerr et al. 2006) of the final system of systems (Neaga et al. 2009; Smith & Oosthuizen 2011; Smith et al. 2012; Sage & Cuppan 2001; Biltgen 2007). In this respect, the analysis can be cost-benefit analysis or cost-effectiveness analysis (Kee 1999). In the literature, fair optimization (Golden 2013; Denda et al. 2000; Ogryczak et al. 2014) and other resource allocation methods (Luss 1999; Harris 1913, Bourdon et al. 2014) have been studied extensively.

Capabilities-based planning is planning under uncertainty to provide capabilities suitable for a wide range of modern-day challenges and circumstances while working within an economic framework that necessitates choice (Davis 2002). One definition of capability is the ability to execute a specified course of action (Biltgen 2007; Biltgen & Mavris 2006). To this end, scenarios are an essential tool for describing different operations and missions. A model for the total capability composed of capability areas is in use in military forces in several countries (JETCD 2009). Military capability areas are defined as statistically independent sets of functionalities. These functionalities

have hierarchies that provide exact definitions of capability areas (JETCD 2009; Smith et al. 2012). The same concept can be used for lower system-level functionalities. Modelling can be conducted in various levels of granularity (Sage & Cuppan 2001). A modeller decides to design an adequate level of the model within the requirements of the problem (Smith et al. 2012).

1.3.2. Related Work: Combat Modelling

A short review of related work of macroscopic combat modelling is presented in this section. An introduction to the history of the subject and more references can be found in (Hartley 2001) or in (Hausken 2000). Most of the macroscopic combat models are deterministic (Bracken et al. 1995). We discuss the development of both deterministic and stochastic combat models.

Combat models can be static or dynamic, that is, time-independent or time-dependent. The first static models were based on numbers of troops and weapons and their appropriate weighting factors. Comparing the weighted sum of the two opposing sides predicts the winner of a battle or a war. T. N. Dupuy (Dupuy 1985) developed the static Quantified Judgment Model (QJM), which takes into account, in addition to weaponry and force sizes, human factors such as leadership, troop morale and operative factors like terrain and weather. Parameters of the model have been calculated with the help of theoretical and empirical results. In QJM, a power potential value for both sides is calculated, and the ratio of the two values predicts the victor.

The most common and the most studied family of combat models are the deterministic Lanchester equations (Lanchester 1914; Osipov 1915). The Lanchester differential equations describe the force sizes in the battle as a function of time. Different battles have been described with quadratic, linear and logarithmic Lanchester differential equations. The equations model the attrition of the two forces.

The quadratic Lanchester equation depicts modern combat with firearms engaging each other directly with aimed shooting from a distance, which can attack multiple targets and can receive fire from multiple directions. The linear Lanchester equation depicts unaimed fire into an enemy-occupied area. The linear law also applies to target acquisition situations. The logarithmic law is typically used to describe the number of casualties of nonfighting participants, such as medical personnel and headquarters staff.

Lanchester's equations have been extended in various ways and empirical studies have been published in (Chen & Chu 2001; Fricker Jr 1998; Hartley 2001; Jaiswal & Nagabhushana 1995; Lucas & Dinges 2004; Salim & Hamid 2001; Speight 2001; Speight 2002). Different forms of the deterministic Lanchester equations give quadratic, linear and logarithmic descriptions for the force sizes. Lanchester equations have been

extensively studied and a goal has been to investigate which of the Lanchester descriptions gives the time-dependent variation of the fighting force size. Historical results of battles do not appear to lead to an unambiguous choice among competing Lanchester equations. Much of the critical effort of the last century has been improving the theory and practice of heterogeneous representation and aggregation and the representation of the different attrition processes.

Also, stochastic versions of the Lanchester equations and probabilistic models based on Markov chains have been proposed. As a model becomes more complex, solving it analytically becomes intractable or computationally expensive. Approximate solutions have been proposed for computationally intensive problems, such as optimal resource allocation and analysis of asymmetric forces like snipers and fighter aircraft (Kim et al. 2017; Lappi et al. 2012; Kress 2020). Most of these models calculate force sizes and attrition rates while probabilities of winning battles and durations of battles are not considered.

Despite its extensive use, the Lanchester formulation suffers from both military (Dupuy 1985) and mathematical (Ancker & Gafarian 1992) inconsistencies. The main critique in (Ancker & Gafarian 1992) is that specific stochastic Lanchester models do not converge to deterministic Lanchester equivalents. The mean outcome of combats is not even defined concerning a deterministic Lanchester model. In this respect, stochastic models, e.g., stochastic Lanchester models, are better suited for combat modelling. Bonder and Farrell (Bonder & Farrell 1970) did the pioneering work on attrition coefficients in heterogeneous target environments. In addition to general inconsistencies of deterministic Lanchester equations, mentioned above, several shortcomings exist in Bonner-Farrell methodology, for example, inconsistent treatment for parallel acquisition of targets (Taylor & Neta 2001). A comprehensive review of combat attrition modelling is presented in (Fowler 1995).

Robert L. Helmbold (Helmbold 1987) and Dean S. Hartley III (Hartley & Helmbold 1995) have studied historical battles and they arrived at an empirical result called the Helmbold relationship. The formula gives a relationship between the attacker and the defender force sizes at the beginning and the end of the battle. The Helmbold relationship does not depend explicitly on any Lanchester attrition laws. The model does not explain the power law of the relationship or the values of the parameters. The data points don't follow the equation exactly because it is an average description of the situation.

Dean S. Harley III has developed his model called Oak Ridge Spreadsheet Combat Model (ORSBM) (Hartley 2001). The model includes force sizes, weapons, human factors, operational variables and environmental variables as independent variables. The most important predicted variables are the force sizes at the end of the battle and the measure of success.

The basic Lanchester laws don't have the rule to end the battle, except when the force size of one side of the battle goes to zero. In (Jaiswal & Nagabhushana 1995) an ending rule has been used, where absolute or proportional force sizes determine the boundary, and expressions for the force sizes at the end of the battle have been derived. Hartley's (Hartley 1995) study shows that the Helmbold relationship may be explained by constraints on the force sizes. The effect of information in military operations has been studied in (Perry 2003). Recently a bottleneck combat model has been introduced (Bongers & Torres 2019).

One of the earliest stochastic models has been published in (Morse & Kimball 1951). In the model, the probabilities of the time-dependent casualties have been calculated for the attacker and the defender. Stochastic war equations have been studied in (Hausken & Moxnes 2002; Hausken & Moxnes 2005) and stochastic variations of Lanchester models have been discussed in (Kingman 2002). There are also agent-based (Hill et al. 2003), fractal-based (Lauren 2002), and cellular automata-based (Moffat et al. 2006) approaches. Salvo models are types of simple micro tactical models, often based on win/lose ideas, and they exist in both deterministic and stochastic (Hughes 1995; Armstrong 2005) forms.

1.3.3. Related Work: Modelling Networked Systems in the Military Context

The resilience of communication networks has been studied widely in the past decades (Colbourn 1987; Ball et al. 1995; Bigdeli 2012). An extensive review article in (Liu et al. 2020) presents network resilience in many other fields.[7] There is a wide range of literature considering algorithms for computing network resilience of connections in general network topology (Ball et al. 1995).

The review article of systems defence and attack models by Kjell Hausken and Gregory Levitin classifies 129 published papers according to the system structure, defence measures, and attack tactics and circumstances (Hausken & Levitin 2012). Failures of operating elements in a system have been investigated in various configurations, e.g. (Levitin et al. 2020).

Using the Markov matrix, or stochastic matrix, formalism is a standard method used to describe the transitions of a Markov chain (Reese 1971; Lindquist et al. 2017). A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event (Karlin & Taylor 1975).

[7] The author of this study has contributions in modelling network resilience considering not only the physical communication but also services provided by application severs located in the network structure (Kuikka 2019b; Kuikka & Syrjänen 2019; Nikkarila & Kuikka 2020; Kuikka & Peltotalo 2020). Information systems and services are important from the viewpoint of this study because information plays a major role in building military capabilities, e.g. situational awareness. Social network analysis (Kuikka 2018a; Kuikka 2021) has also military applications (Kuikka & Latikka 2018; Kuikka 2018b; Kuikka 2019c; Kuikka & Nikkarila 2019).

Robustness and vulnerability design for autonomic management has been studied in (Bigdeli 2012) where network robustness measures of network criticality and algebraic connectivity have been presented in the context of network management. Software Defined Network (SDN) controller placement, based on network measures of network criticality (Bigdeli et al. 2009) and algebraic connectivity (Fiedler 1973; Bigdeli et al. 2009), to achieve robustness against targeted attacks have been investigated in (Alenazi 2018).

Social network analysis in organisational studies has been utilised in (Sozen et al. 2009; Hoppe & Reinelt 2010). Leadership in military organisations (Goodwin & Blacksmith 2018; ATP 2016) has been studied in (Kuikka & Nikkarila 2019; McChrystal 2015).

1.4. Research Gaps

Military concepts have been studied with different methods, mathematical and non-mathematical. One of the non-mathematical methods that are related to this study is conceptual analysis. However, we have delimited conceptual analysis outside this work, and instead, build on existing definitions of military concepts and their relationships.

Primary Research Gap
Our primary goal is to form a common mathematical methodology that is capable of describing quantitatively a wide range of central military concepts. As we have selected military capabilities, combat outcomes and networked systems as our representative military concepts, a natural choice is using probability theory. There are not many other mathematical theories that we can think of which meet our requirements.

The first problem that has not been answered or demonstrated in existing research is how to develop mathematical models or how to apply existing mathematical theories within a common mathematical framework that can quantitatively describe concepts like military capabilities, combat effectiveness or outcomes and networked systems. Also, we have research gaps inside these three sub-areas.

Research Gap 1
In the existing literature, no quantitative mathematical methods have been presented that can describe military capabilities on different hierarchical levels. Partial solutions exist for the number of troops and armament but they don't allow comparing of different capabilities or adversary forces.

Research Gap 2
In the existing literature, no probabilistic methods have been presented that describe combat effectiveness, combat outcome and expected combat duration in terms of analytical closed-form expressions.

Research Gap 3
In the existing literature, no quantitative mathematical methods have been presented for describing networked systems that use a common methodology for military capabilities and operability of networked systems and services. As a specific case, we find that there are no mathematical methods for calculating optimal sensor placement in a network structure that takes into account sensor functionalities and protection requirements.

In summary, under the main question of how to describe central military concepts with a common analytic mathematical methodology, we have identified three research gaps in the modelling of the selected central military concepts. As we have decided to use probability theory as our methodological approach, these research gaps can be formulated as follows.

1.5. Research Questions and Delimitations

The main research questions in this study revolve around mathematical modelling of military concepts and applying probabilistic theories in the military domain. We aim to develop mathematical models for describing central and interrelated high-level military concepts. Three concepts, that fulfil these requirements, are selected as representative target areas of modelling: combats, military capabilities and networked systems. We present a set of mathematical methods that can be applied to these areas. We provide practical examples of applications to show that the methods are useful and they give reasonable and consistent results.

Table 1.1 summarises highlights and some interdependencies of the three target areas of modelling. Highlights are presented on the diagonal elements of the matrix and they are discussed in more detail in Sections 2, 3 and 4. The off-diagonal elements show examples of how the three concepts can contribute to each other in the context of this study. For example, interconnections between battles, forces, armament and location information could be modelled by the methods of networked systems. Interrelations between the three concepts are bi-directional and further dependencies can be identified when sub-structures of capabilities, combats and systems are considered. This study covers some of the interdependencies, but many of the applications have not yet been discovered.

The purpose of using the methods is to help military planning and to support decision-making processes through clarifying military concepts (explaining) and providing quantified results (predicting) within the applicability domains of the models.[8] It is

[8] (Low 1995): The absence of prediction capability has ramifications for the model applications; specifically, those concerned with estimating force levels and support requirements. These estimates, to the extent they are based on models, are closely tied to absolute rather than relative model results as appropriate force parameters are varied in the analysis. It is unfortunate that such absolute measures remain beyond our grasp. Nevertheless, we retain a lot of the utility inherent in applying models of combat to military problems even though we are

noteworthy that, even if a model turns out to be too simple or unsuitable for the analysis needs, in a particular problem, the process of evaluating and discussing different models can help clarify the common understanding of concepts and objectives. We try to keep the methods on a fundamental level, while at the same time, introducing means of expanding the models for more detailed and granular systems.

Research Questions

Our goal is to find a common framework or methodology to be used in analysing and modelling high-level concepts in the military domain. Both purposes of modelling, explaining concepts and predicting outcomes, are considered equally important in supporting military planning and decision-making processes. We develop a cohesive and extendable set of mathematical models that can be used to study central concepts in this area. To this end, we select three representative concepts: military capabilities, combat outcomes and networked systems. We use quantitative methods that enable developing models with a common methodological basis. An important consequence of this approach is that mathematical quantities, both explanatory and response variables, can be expressed by similar terms and compared with each other.

Table 1.1. Highlights of the models and some of their interdependencies.

	Capability	**Combat Modelling**	**Networked Systems**
Capability	Capabilities, system capabilities and sub-capabilities can be modelled with the proposed methods	Engagement is one of the high-level capability areas	Network-centric capabilities constitute one of the high-level capability areas
Combat Modelling	Military capability is the main factor determining the probability of victory and expected duration of a battle	Combat modelling can include force sizes or forces sizes and technical systems (e.g. arms and weaponry)	Interconnections between battles, forces, armament and location information can be modelled by the methods of networked systems
Networked Systems	Military capability can be used to attack adversary network structures and military capability can be used to protect own networks and networked services	Networked systems can be modelled as a part of military force or environmental factors in combat modelling	The resilience of networked systems, the utility of services and organisational structures can be modelled with the proposed methods

forced (by the lack of scientific knowledge and methodology) to deal with the relative outcome measures. In so doing, we underscore a tutorial attribute of modeling: that of providing insights into problems of great complexity. This attribute can be put to use when applying models to areas of military education, training, and systems development. Here it would seem that deterministic models (hopefully accredited or validated in some manner) could, in certain circumstances, provide useful relative results that would signal preferred operational procedures or point to worthy directions for technical development. This constitutes a use of modeling more prescriptive than predictive and also applies to state-of-the-art stochastic models, imperfect though they may be.

The primary research question of this study is the following:

How to describe central military concepts through novel quantitative models to support planning and allocation of networked system capabilities?

To answer the primary research question, we set the following sub-questions:

1 What novel mathematical models quantify military capabilities in different hierarchical levels: capability areas, system capabilities and functional capabilities?
2 How to calculate the probability of combat outcome and duration of battle with a model?
3 What novel or existing probabilistic models describe the operability and functionality of networked systems?

We divide research sub-question 1 into the following objectives:

1.1. Develop a model that uses capability values (or their changes) on a high functional level as predictor variables and has system capability values as dependent variables.
1.2. Develop a model for describing the dependency of military capabilities on technological development.
1.3. Develop a model for calculating the optimal number of system units within a predefined budgetary constraint.

We divide research sub-question 2 into the following objectives:

2.1. Derive a novel combat equation predicting individual battle outcomes by the methods of stochastic analysis. Develop a novel model for predicting the probability of victory for individual battles.
2.2. Develop a novel model for predicting the expected duration of individual battles.

We divide research sub-question 3 into the following objectives:

3.1. Present mathematical methods for describing networked systems by using the Markov matrix formalism of probability theory.
3.2. Develop a methodology for planning sensor system configurations in network topology and models for calculating the optimal sensor placement that considers the number of sensors, the resilience of sensor locations and optimal access to controlled or monitored nodes in the network.

Delimitations

We exclude the comparative and conceptual analyses of different definitions and specifications of capabilities and their sub-structures. The reason for this is that the mathematical models are intended to be applicable for different capability schemes that implement some general requirements. We provide mandatory requirements for

capability schemes that should be fulfilled for the capabilities and sub-capabilities to be investigated with the models of this study. After all, a small sub-set of well-defined concepts is sufficient for many practical cases.

1.6. Research Methods

In this section, we present the research methods of this study. We make use of methods generally used for an operations research study (Jaiswal 1997; Gass 1983). Fig. 1.2 is a schematic diagram representing these various steps.

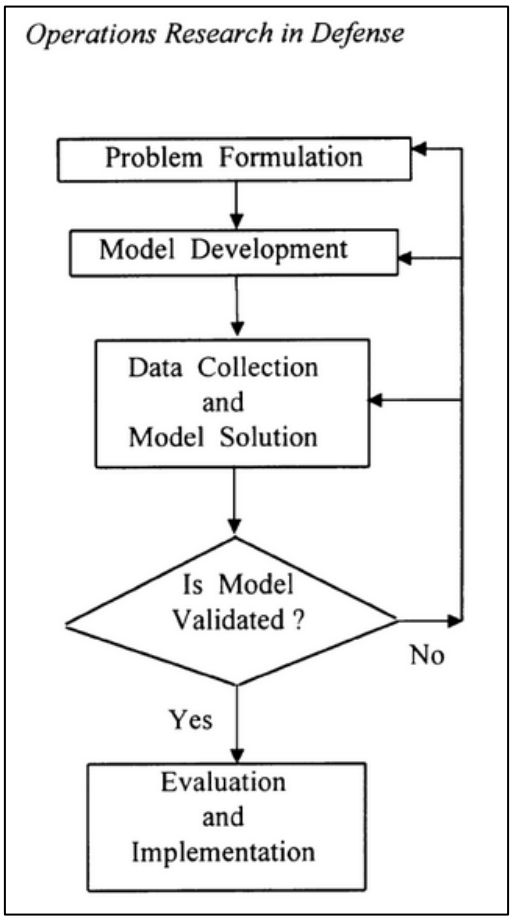

Fig. 1.2. Steps of an operations research study, figure adapted from (Jaiswal 1997).

Problem Formulation

As we aim to develop quantitative models, we use probabilistic methods to describe three concepts of military capabilities, combat outcomes and networked defence systems. The common methodology enables combining these models across these areas when solving problems in operations research, such as planning capabilities of networked systems to enhance combat effectiveness. Because the three concepts are related, we have goals that we summarise as follows: the proposed models should be extendable to describe also other applications in the military domain. Also, the models should be extendable to a desired structural level of detail.

Model Development

Modelling military capabilities with probabilistic methods enables us to develop a common methodology for various concepts of military capabilities. Armament, will and troops can be described quantitatively with probabilistic models. In this study, one objective of operations research is to model capability areas and their sub-capabilities. The general concept of 'system of systems' is used to describe how various systems generate services that can be identified with sub-capabilities in the conceptual capability model (Kosola 2013). Necessary and alternative sub-capabilities in the scenario are modelled with well-known probabilistic concepts like mutually exclusive events.

In the second part of this study, we model combat effectiveness by using probabilistic methods. Stochastic analysis (Florescu 2015) and Itô formalism (Karlin & Taylor 1975; Florescu 2015) is used to model attrition processes of two adversary forces. An analytic closed-form combat equation is derived assuming that the force sizes of the two sides obey geometric Brownian motions (Florescu 2015). Decision boundaries are used as stopping rules for combats. We have also derived analytic closed-form expressions for combat durations by using slightly different definitions for stopping rules and stochastic processes describing the combat.

The third objective of this study is to develop probabilistic models for describing networked systems. Networked systems are ubiquitous in all levels of military operations, and for this reason, it is an essential building block in developing mathematical models for describing a sufficiently wide range of military concepts. Two specific cases are included in modelling networked services: to model operability with the theory of Markov processes and to optimise sensor placement in a network structure. Both aspects of technical network resilience, and utility of information services, can be considered within the models.

Data Collection and Model Solution

Due to the broad scope of data collection and data analysis generally, we only discuss the data collection and the use of the models in the specific context of this study. The main purpose of using representative data with the models has been to demonstrate the validity of the models. Historical combat data has been collected in (Hartley 2001) which is the largest available data set. This is appropriate for model validation but insufficient to extrapolate the range of validity to modern, and more complicated, combats. More detailed modelling with additional features, such as incorporating information networks or other capabilities, may be necessary. Moreover, a general problem with military systems is that no empirical experiments are possible (Jaiswal 1997).

We conducted a questionnaire among ten students in military technology. Possible problems with the questionnaire are the small number of participants, the level of expertise among the participants and biases based on human judgment (P II). Despite these problems, answers to the questionnaire questions were consistent and variations

between the numerical values have interpretations in real situations (Suojanen et al. 2015; Suojanen et al 2014; P I). Note that the models in (P I, P II, P III) can be used for interpolation and extrapolation outside the original data points of 1 year, 10 years and 20 years. However, while extrapolation is technically valid, assuming explanatory variables have reasonable values, they are limited to a time interval of between 20 and 40 years. Estimations of the parameters are based on questionnaire data and the time horizon for extrapolating these results may be even shorter.

Two questionnaires have been conducted in (P I, P II, P III) as explained in (Suojanen et al. 2014). Historic combat data is used in (P IV, P V) from (Hartley 2001). Validity of the applications in (P VI) has been conducted by using typical model parameter values, where the method itself is based on standard Markov processes. The network structure of the Sprint infrastructure network in the US, used as input data in (P VII), has been retrieved from a public database.

Model Validation and Evaluation

In this section, we discuss the validation of the models in this study. Model validation tests the agreement between the behaviour of the model and the real-world system being modelled. If the results of the system operation are available, we can statistically compare model results with those obtained from system operation. Military system studies, however, suffer from the lack of operational data and therefore in most cases, the model validation gets limited to the perception of the military experts (Jaiswal 1997). The problem of model validation has been extensively discussed in the operations research context (Gass 1983). Three aspects of validation are mentioned in (Jaiswal 1997, Gass 1983): face validity or expert opinion, variable – parameter validity or sensibility analysis and hypothesis validity.

Saul I. Gass: "A mathematical model of an ongoing physical system must be able to replicate the past to be accepted as a decision aid. But what of simulation and policy analysis models of non-existent systems or for future-oriented situations in which the past is not a good predictor of the future? For first-time futuristic models, validity is superseded by the concept of model credibility, as defined by the decision-maker." For a first-time model, a model of a non-existing system or one that makes assumptions about the possible future states of the system, the following concepts apply - face validity or expert opinion, variable - parameter validity or sensitivity analysis, and hypothesis analysis. (Gass 1983)

Face validity or expert opinion answers the following question: "When the model is demonstrated to experts who are aware of the system being modelled, do they feel satisfied with the behaviour of the model, that is, is the model credible? Variable – parameter validity or sensibility analysis answers the following question: "If an analysis is carried out, do the changes in model output due to changes in variables and parameters occur following the real-world system?" Hypothesis analysis answers the following questions: "Do pairwise or higher-level relationships correspond to similar

relationships in the real world? Do the subsystem models interact realistically?" (Jaiswal 1997)

Evaluation and implementation lie on the decision-maker. The usability of the models of this study is demonstrated in example applications but ultimately the evaluation should be conducted in the context of actual implementations.

Reliability in Model Predictions

Methods for dealing with uncertainty in quantitative modelling have been classified in (Abrahamsson 2002). Obtaining acceptable reliability in model predictions requires managing uncertainties related to:

1. *specifying the correct problem* or scenario the model is intended to address (Otherwise the model 'produces correct results for the wrong problem'),
2. *formulating a conceptual model of the problem* taking all relevant entities, processes and interactions into consideration, whilst omitting the rest,
3. *formulating a computational model* with a suitable resolution for the intended use of the model,
4. *estimates of the input parameters* to the computational model,
5. *calculation and documentation of results*.

The previous list originally presented in (Abrahamsson 2002) is adapted from (Andersson 2018). In the classification of (Gass 1983), reliability in model predictions is mainly included in the last step 'evaluation and implementation' in an operations research study in Fig. 1.2. In this study, we regard 'reliability in model predictions' as referring to the implementation and operational use of a model and validity of the model as referring to the model itself. As noted in (Andersson 2018) the operations research approach prescribes involving the end-user organisation when working out relevant scenarios and concepts of operation. As a consequence, estimating the reliability of a model is a cooperative effort of the modeller and end-users of the model.

2.

CAPABILITY MODELLING

Military capability is a powerful concept in the planning and modelling of military strength. Military capability is a composition of different existing and potential strengths. The concept of capability has been used to express the level of will, amount of troops and quantity of armament. Combat capability is a generalised characteristic of quantity and quality of forces and assets. Measuring military capability and the difficulty of quantifying military capability in a single definite measure, have been discussed in (GAO 1985; Craig et al. 1991; Biltgen & Mavris 2006; Saperstein 1994; Jürgen 2003). In this study, we propose methods for modelling military capabilities with probabilistic terms. Here, a capability is defined as the probability of successful a mission or operation in a specified scenario (Suojanen et al. 2015; Tugrul et al. 2013).

2.1. System of Systems

Military capability is a widely used concept in planning and developing a nations military forces (NATO 2019; DoD 2020; Tellis et al. 2000). We propose a general method for modelling military capabilities with probabilistic methods. Capability is defined as the probability of mission or operation success in a specified scenario. System capability is defined as the probability of a functioning system. This definition can be interpreted as a quantitative refinement of the original definition in (Kosola 2013). The definition based on probabilities is consistent with the earlier concept and also in line with many other corresponding descriptions of military capability in the literature (NATO 2019; DoD 2020). As a prerequisite, the scenario (Suojanen et al. 2015) and the meaning of successfulness in the operation have to be described at a desired level of exactness. These descriptions should follow the same level of fidelity and precision as the model itself which can be at a high level for capability areas or at a more granular level of systems and functionalities.

The model in (P I) relates the capability of a system to capability areas of situational awareness, protection and engagement. We demonstrate the methodology with questionnaire data (Suojanen et al. 2014; Horne & Schwierz 2008) and two systems, the satellites and the UAVs in three scenarios and three time-frames. The respondents of the questionnaire were instructed to provide their evaluations as changes in comparison to the present capability values for the three capability areas. Changes in capability areas are positive, as expected because implementing new systems should improve capabilities. In (P I), capability changes for situational awareness are presented in 1 year, 10 year and 20-year time horizons for the three scenarios. Corresponding results for protection and engagement capability areas are presented in (Kuikka et al. 2015).

Only general specifications of systems and scenarios were available for the group of ten respondents of the questionnaire. Discussions and searching for additional information were recommended. Still, respondents may have different opinions about the needed number and type of satellites and UAVs. Some UAVs may have firing capabilities and some UAVs have only communication or situational awareness capabilities. Changes in the values for capability area levels are calculated as averages from individual answers and then used in the analysis. In this way, a common average understanding of the group is formed. (Suojanen et al. 2014, Suojanen et al. 2015)

The questionnaire data contains changes in capability values for all mentioned capability areas. From the questionnaire data, we can calculate individual system capability values by reverse engineering. Equations of the model determine relations between the capability area values and the system level values. In this way, the model enables solving the unknown system capability values. This is one of the advantages of the methodology.

Systems have different functionalities in protection, awareness and engagement capability areas. Alternative defence systems and their combinations can be compared to maximise the overall capability. Constraints of available resources can also be considered. This kind of analysis is presented in (P II). Typically, multiple systems can contribute to the same capabilities and systems are alternatives to perform the same mission functionalities. On the other hand, complementary systems can be used simultaneously to maximise the impact. Technically, alternative systems are modelled as parallel capabilities and necessary systems are modelled as serial capabilities. Alternative systems provide similar, or exchangeable, functionalities possibly with different capability values. This requirement of producing parallel capabilities may lead to constructing a more detailed model by dividing functionalities into sub-functionalities. (P I)

Respondents answered questions about changes in capability area values after deploying satellites or UAVs or both of them (see Fig. 2.1). This information is used as input data for the equations in the model. The model has four different systems: satellites, UAVs, parallel systems and auxiliary systems. Parallel and auxiliary systems may consist of multiple individual systems or manual processes. Because the model includes parallel and auxiliary systems, interesting conclusions can be made that are not obvious from the questionnaire results. The results show that parallel system capability values decrease in all scenarios as a function of time. This has a logical explanation as the new systems can take over functionalities and capabilities from existing systems. In some scenarios system capability values of auxiliary systems increase as a function of time. Our model predicts that these changes account for many changes in capability areas values instead of satellites or UAVs. (P I)

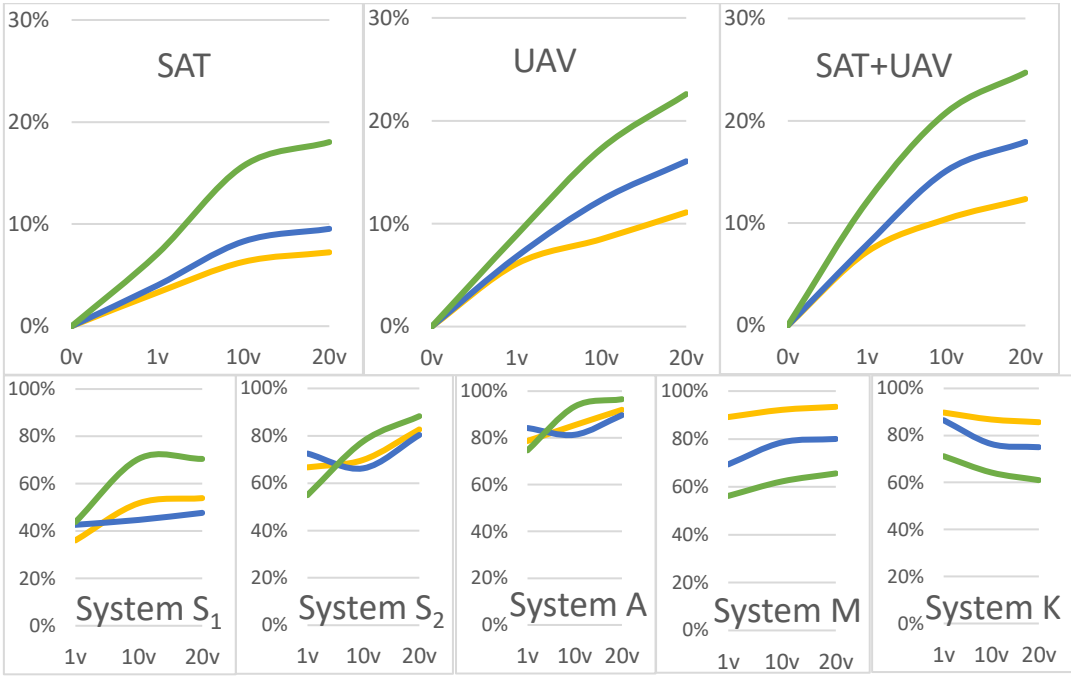

Fig. 2.1. Mapping of changes in capability values of capability areas (top) to system capabilities (bottom). Capability areas results are from the questionnaire and system-level capabilities are calculated from the model. The figure is adapted from (Kuikka et al. 2015). Curves are for situational awareness and three scenarios (Scenario 1: yellow, Scenario 2: blue, Scenario 3: green). System capabilities (bottom) are for systems S_1 (satellites) S_2 (UAVs), A (satellites and UAVs), M (necessary) and K (alternative).

Complex relations between different system capabilities and capability areas are modelled with the help of probability theory (Florescu 2015) as a union of non-exclusive events of successful operations. The use of probability theory simplified the modelling task significantly. The most important advantage of the approach is that systems are modelled only once in the framework of a system of systems and there is no need to model all pairs of systems like in the analytical hierarchy process (Saaty 1990). In our model (P I), adding a system to the model or removing a system from the model does not require recalculation of the model parameter values of the other systems of the model. In this particular case, adding a third system, such as fighters, does not change the parameter values of UAVs or satellites if their functionalities remain the same as before. However, structural changes in equipment or operational use can result in changes in the model or its parameter values.

Effects on capability areas or total capability can be calculated more directly than in most other methods. This feature makes comparing capabilities of different system configurations easier. Also, the model can be useful in planning alternative development and acquisition projects where accurate or itemised technical information may not be available. More detailed models may not be used because not enough reliable input information is available.

Different specifications of capability areas can be used within the model and desire levels of granularity can be included in the model. The idea of parallel and auxiliary systems can be used as a first model in cases, where detailed information about the system and its functionalities are not known, to serve as a tool for understanding the main characteristics of the systems modelled by the method. Also, the first version of the model (P I) could be refined by including more functionalities or sub-functionalities in the next iteration of the method. After all, which kind of information is available and the goals of the modelling, determine which methodologies and models are most useful in each case.

The method allows the modelling of a limited set of systems or functionalities. It is possible to model high-level capabilities with only one level of hierarchy describing a system-level below the main capability areas. This can be the first phase in the process of selecting more systems to be included in the model and gaining an understanding of the many dimensions of the problem. It may be easier for domain experts to evaluate capability developments on a high capability area level (see Fig. 2.1) than on a system-level because of complex interrelationships between different systems. In addition to the deployment of new systems and general technological development, tactical and other improvements in the operational environment may take place at the same time. In many cases, the main focus is on capability area probabilities or the probability of a successful operation while original causes of capability changes may only be of secondary interest.

One of the main applications of the modelling approach (P I) is that it can be used to discover underlying causes for high-level capability changes that are impossible or difficult to find out in raw questionnaire results. The idea of the questionnaire was to evaluate capability values on the capability area level (top row in Fig. 2.1) instead of the individual systems capability values. In the second phase system capability values can be calculated by reverse engineering from high-level capabilities. This is possible when alternative (parallel) systems are evaluated both individually and pairwise. In this respect, the approach can be characterised as a top-down method. The model is a tool to decompose the capability area information into system and sub-system components. Numerical examples of the analysis are provided in (P I). More interpretations of the results of the system capabilities are presented in (Kuikka et al. 2015) where the examples demonstrate results with three scenarios for different capability areas and systems.

2.2. Technological Development

Our goal is to get a better understanding of the dependencies behind the capability changes in high-level capability areas. The focus is on future developments of military capabilities in the context of technology forecasting but the models are suitable also

for investigating present state capabilities and applications outside of the military context. The methodologies include operational analysis, systems' modelling, probabilistic mathematical methods, statistical evaluations, conceptual modelling and the use of questionnaires as input data.

Disruptive technologies can create new sources of military capabilities but can also degrade competence. In the hands of an adversary force, advanced technologies and their use may pose a threat. Uncertainties of technological development that characterise emerging technologies mean that the military cannot know which technologies mature to have profound impacts. Most emerging technologies represent incremental improvements and enhance the competencies of the military. An emerging technology that undermines existing training, equipment and doctrine will have more impact on the military than one that complements or enhances existing military capabilities. (James 2013)

We focus on interdependencies of technologies and impacts of technologies on military capabilities by taking into account the influences of multiple technologies on system capabilities and top-level capability areas (Kuikka et al. 2015). In (Kuikka et al. 2015), we provide a literature review related to the concepts and models of our work. The review serves also as an introduction to the concepts used in the research of innovations and technological development. In the literature review, we discussed the similarities and differences of our method and corresponding models in the literature. Different mathematical methods have been used when multiple interdependent variables are affecting the forecast. Examples are principal component analysis (Windrum et al. 2009), simulation methods and time series methods (de Gooijer & Hyndman 2006). Typically these methods need more data than is available in our study. Judgmental forecasting methods incorporate intuitive judgments, opinions and subjective probability estimates. This characterization is also valid in our study. Modelling capabilities is usually based on scoring or other measures with no quantitative interpretation.

Martino (Martino 1993; Kim 2012) proposed a scoring model for rating technology:

$$\hat{S} = \frac{A^a B^b (cC + dD + eE)^z (fF + gG)^y (1 + hH)^x}{(iI + jJ)^w (1 + kK)^v} \tag{1}$$

where $c + d + e = 1, f + g = 1, i + j = 1, a + b + z + y + x = 1$ and $w + v = 1$. Variables A and B in the model are overriding factors and $\{C, D, E\}$, $\{F, G\}$ and $\{I, J\}$ are exchangeable factors within brackets. I, J and K are costs or undesirable factors. The factors $(1 + hH)$ and $(1 + kK)$ represent special cases that must stand alone but cannot be traded off with any other factors. Moreover, they may not always be present. In the equation h and k are constants. Martino's model can be compared with our model as the serial system X_m can be considered a combination

of overriding factors (Kuikka et al. 2015). Parallel systems correspond to the exchangeable factors in Martino's formulation. Cost or undesirable factors are not examined explicitly in our model, their effects are included implicitly in the input capability values. In our method, systems of systems are modelled with serial and parallel systems and extra factors like $(1 + hH)$ are not needed. (Martino 1993; Kim 2012)

The main concepts of describing military strength from the technological point of view are capability, systems and technologies. The capability has been defined as the ability to achieve a specified wartime objective, e.g. win a war or battle, or destroy a target (NATO 2019, DoD 2020). This definition is compatible with the probabilistic definition used in this study (P I). Systems are characterised as having structure, behaviour and interconnectivity. Technology is the collection of techniques, skills, methods and processes used in the production of goods or services (Windrum et al. 2009). It is the state of knowledge of how to combine resources to produce new products, to solve problems, fulfil needs, or satisfy desires. (James 2013)

Three levels of abstraction are needed for a model to describe the effects of technologies on military capabilities. These levels are capability areas, systems and technologies (P II). We present a method to model interdependencies between technologies. Modelling interdependencies between technologies is one step in building a comprehensive quantitative model for military capabilities. As in (P I), a capability is defined as the probability of a successful task or operation or proper functioning of a system.

On the system level, the model makes use of the system of systems principles. Satellites and UAVs have been assembled in parallel or series with other systems. This idea can be compared with the ideas in Martino's more heuristic model (Martino 1993). Other models, like the analytical hierarchy process (AHP) (Inman et al. 2006), require all the weights to be recalculated after a new system or a task is added into the calculation. In our method, if no significant secondary effects exist, a new system can be added in serial or in parallel to an existing set of systems with a limited amount of recalculations (Kuikka 2019).
To obtain numerical data for demonstrating the model, we conducted a questionnaire to a small group of defence technology researchers. Questionnaire data is used as an input for modelling relationships between operational tasks, systems, technologies and capability areas. The functional form of the model gives an approximation for calculating the effects of different technological developments on capabilities. The model takes into account the recurring effects of multiple technology areas. One consequence of this is that the method using the interdependency of technologies is a tool to examine the effects of new developments with updated data, for example, more effective sensors become available. (P II)

Interdependencies between seven representative technologies were inquired in the questionnaire. They are sensor, material, communications, stealth, energy source,

manufacturing and autonomous technologies. Five different operational tasks are used in the evaluation: surveillance, communications, engagement, logistics and deception. We show how to evaluate interdependencies between technology areas. For example, material technology has a significant effect on sensor technology. Meaning that progress in the first area implies progress in the latter area. Usually, interdependencies are not symmetric, for example, sensor technology depends significantly on material technology, while the dependency in the opposite direction is weaker. Our model can be compared with the stochastic correlation matrix of statistics (Florescu 2015). The correlation matrix is symmetric while the model proposed here allows different values for dependencies depending on the direction.

The functional form of the model (P II) describing task capabilities for the seven tasks is

$$C_j = \frac{1}{7}\sum_{i=1}^{7} A_{i,j} \prod_{k=1}^{7} \left(1 + D_{i,k}(\Delta_k - 1)\right), j = 1, \dots, 5, \tag{2}$$

where matrix elements $A_{i,j}$ describe the development of technology area i on task capability j. Matrix D gives the interdependency between technology areas. Numerical values of the matrix elements of A and D are provided in (P II). Parameter Δ_k is used to model slower or faster technical development concerning the original input data. This enables calculating the effects of different linear developments in technology areas without repeating the questionnaire or changing the model parameters. For example, $\Delta_k = 0$ means no development, $\Delta_k = 1$ means original evaluation, and $\Delta_k = 2$ means 100 % more development than in the original input data.

The method is useful in illustrating the general impacts of various technological areas and their interdependencies. The results in (P II) indicate that material technology has the most widespread influence on the other six technology areas. UAV material and sensor technology have a great influence on surveillance because surveillance task is improved by radars and other sensors. Logistics task is not sensitive on the development of sensor technology in satellites. Modest development in stealth technology does not affect greatly the performance of any of the UAV tasks and affects weakly the performance of the satellite tasks. In general, slow development in any technology area affects more the development of UAV than satellite capabilities.

2.3. Mathematical Methods

We study two related topics – capability modelling and analytical growth curves applied to capability values (Kuikka 2019a). We also present a general method to calculate contributions of more than two additional parallel systems (P I, P II). The procedure is used iteratively for any number of capabilities with additional systems deployed incrementally on top of serial and parallel infrastructure capabilities and existing alternative capabilities.

Probabilistic modelling methods and analysis are also extended with the use of analytical S-shaped growth curves for capability areas and system capabilities (Kuikka 2019a). We illustrate how the derived results, that is, system capabilities and derivatives of system capabilities, can be used to discover different characteristics and trends, which are difficult to discover from high-level aggregate capability areas data. After detecting these main trends, a more detailed investigation can be directed to the most important findings in next the steps of the analysis. Also, mathematical formulas for evaluating errors in derived quantities are derived in (Kuikka 2019a). This enables studying error propagation from uncertainties in estimated capability changes to calculated system capability values.

Here, we use the general concept of capability for different levels of capabilities. A capability can have a structure composed of parallel and serial sub-capabilities. Higher-level capabilities are constructed from these building blocks of lower-level structures. Serial capabilities are easier to deal with because the capability value on a higher level is the product of serial sub-capability values. Next, we present how parallel capabilities can be computed. We denote existing capabilities by M and K. M consists of the necessary capabilities required for the operations. K consists of a system of systems having similar functionalities with d new systems added to the system of M and K. The arrangement of the system of systems after adding the d new systems is shown in Fig. 2.2. System capabilities of systems $M, K, S_i, i = 1, \dots, d$ are denoted by $X_m, X_k, X_i, i = 1, \dots, d$.

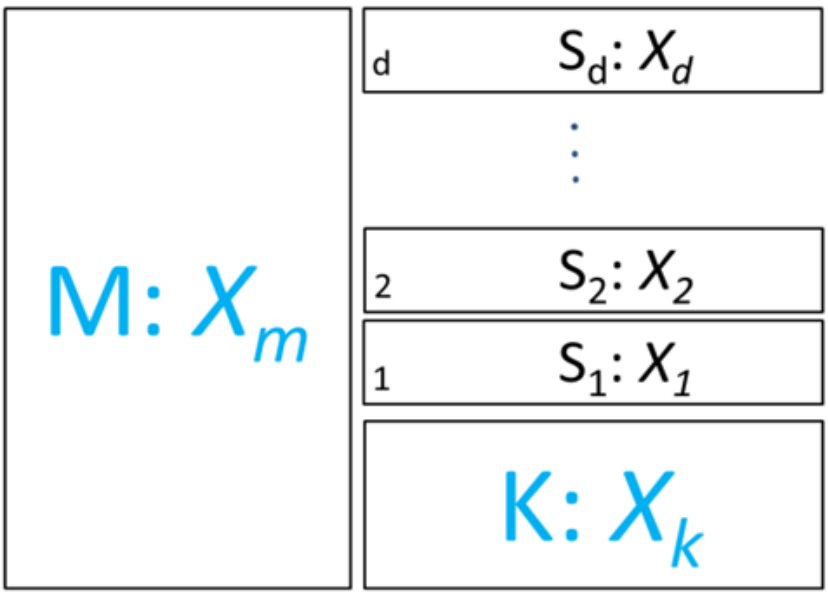

Fig. 2.2. Systems M and K with additional systems $S_i, i = 1, \dots d$ parallel to system K. System capabilities are denoted by X_m, X_k and $X_i, i = 1, \dots d$.

The capability value of the composed system of systems M and K in the operation is assumed to be p_0. Because the two systems are serial, we have $p_0 = X_m X_k$. First, we add a new system with capability value X_1 in parallel with K. The capability of these

three systems K, M and X_1 is

$$p_0 + p_1 = X_m(X_1 + X_k - X_1X_k) = X_mX_1 + p_0 - X_1p_0. \tag{3}$$

X_1 can be solved as

$$X_1 = \frac{p_1}{X_m - p_0}. \tag{4}$$

Next, we add the second system with capability value X_2. The capability of these four systems K, M, X_1, X_2 is

$$\begin{aligned} p_0 + p_{12} &= X_m(X_1 + X_2 + X_k - X_1X_2 - X_1X_k - X_2X_k + X_1X_2X_k) \\ &= p_0 + (X_m - p_0)A_{12}, \end{aligned} \tag{5}$$

where

$$A_{12} = 1 - (1 - X_1)(1 - X_2). \tag{6}$$

The value of X_2 can be solved because the value of X_1 is known:

$$X_2 = \frac{1}{(1 - X_1)}\left(\frac{p_{12}}{X_m - p_0} - X_1\right). \tag{7}$$

In general, the following iterative formula holds (Kuikka 2019a):

$$X_i = \frac{1}{(1 - A_{1\ldots i-1})}\left(\frac{p_{1\ldots i}}{X_m - p_0} - A_{1\ldots i-1}\right), \tag{8}$$

where

$$A_{1\ldots i-1} = 1 - \prod_{j=1}^{i-1}\left(1 - X_j\right), \text{ where } i = 2, \ldots, d \text{ and } A_0 = 0, A_1 = X_1. \tag{9}$$

The procedure describes the building of a combination of systems when new parallel systems are added, one by one, to the existing system of systems. The quantity $p_{1\ldots i}$ is the added value of capability on the capability area level when the system S_i is added parallel to the systems $S_i, i = 1, \ldots i - 1, i \leq d$ (see Fig. 2.2). Note that $p_{1\ldots i}$ is the value added to the system of systems $M, K, S_i, i = 1, \ldots, i - 1$. The quantity in $A_{1\ldots i-1}$ in Eq. (9) is the system capability value of the system of systems $S_i = 1, \ldots, i - 1$. The system capability values of systems M and K do not affect the value of $A_{1\ldots i-1}$. (Kuikka 2019a)

The iterative formula for X_i assumes that X_m is known. This is possible if the capability value of system capability X_m is evaluated separately. Alternatively, if the capability values are known for an increasing set of systems, X_m is solved from the above equations. For example, for two parallel systems the value of $X_m - p_0$ can be solved as

$$X_m - p_0 = \frac{p_1 p_2}{p_1 + p_2 - p_{12}}. \tag{10}$$

This is an example of how unknown capability values can be solved when a sufficient number of system capability values and combined systems' capability values are known. The number of unknown quantities and the number of equations must be equal.

2.4. Protection Capability

In this section, we extend the analysis of the protection capability area in (Kuikka 2019a). We noted that the definition of protection capability is more abstract than the other two capabilities of situational awareness and engagement analysed in (Kuikka 2019a). One consequence of this is that the capability value $X_m = 1.12$ (112 %) for Scenario 1 at time $T = 10$ years. This is not possible because probabilities greater than 100 % are not possible by the definition. We have indicated the value of $X_m = 1.12$ by the red colour in Table 2.1.

There are several possibilities for what went wrong. One possibility is that in the questionnaire the estimated capability changes have high variances because of the complex nature of the concept of protection capability. During the questionnaire, no automatic consistency checking between the evaluated capability changes of satellites, UAVs and combined use of satellites and UAVs were conducted. The second possibility is that the model described in (P I, P II, P III) does not describe the protection capability of satellites and UAVs accurately enough in this particular case. Both of these reasons may contribute to the unphysical probability value in Table 2.1 (and corresponding tables in (Kuikka et al. 2015; Kuikka 2019a).

Next, we analyse in more detail the case of protection capability in Scenario 1 at time $T = 10$ years by making two alternative modifications to the assumptions we had made with the parameters of the model. In this way, we try to show how we can still use the model and possibly make estimates about more realistic modelling results. Because the value of X_m is too high, we can make two changes to reduce the value: reduce the value of the initial capability value x_0 from $x_0 = 0.7$ to $x_0 = 0.58$ or less in Table 2.1, or reduce the value of parameter α from $\alpha = 1.0$ to $\alpha = 0.85$ or less (see Equations (7) – (13) in (P III)). We have indicated these parameter values by the blue colour and the changed results of the model by green the colour in Table 2.1.

With the value of the initial capability value of $x_0 = 0.58$ the system capability values of X_1 and X_2 remain the same. This is a general property of the model as shown in Section 2.4. With the parameter value of $\alpha = 0.85$ the system protection capability values of satellites $X_1 = 0.15$ and UAVs $X_2 = 0.21$ are slightly higher than the corresponding values of $X_1 = 0.13$ and $X_2 = 0.18$ calculated with the parameter value of $\alpha = 1.0$.

The values of quantities X_k, A, X and P_{prot} change in both cases of $x_0 = 0.58$ and $\alpha = 0.85$. With $x_0 = 0.58$ the values of $X_k = 0.58\ (0.63)$ and $XA - x_0 = 0.70\ (0.82)$ are lower than the original values. The original values are shown in parentheses. With $\alpha = 0.85$ $X_k = 0.70\ (0.63)$ and $A = 0.34\ (0.29)$ are higher and $X = 0.30\ (0.42)$ is lower.

We conclude that the parameterisation with $\alpha = 0.85$ agrees better with the original questionnaire answers because the value of protection capability in Scenario 1 at time $T = 10$ is $P_{prot} = 0.82$, which is closer to the value of protection capability $P_{prot} = 0.80\ (\alpha = 0.85)$ than the value of $P_{prot} = 0.70\ (x_0 = 0.58)$. The interpretation for the parameter value of $\alpha = 0.85$ is that satellites and UAVs in Scenario 1 at time $T = 10$ are used together in a way that enhances the capability of the combined system from $A = 0.29$ to $A = 0.35$ (see Table 2.1). The value of the protection capability P_{prot} remains almost the same because the value of $X = X_m - x_0$ has decreased from $X = 0.42$ to X=0.30.

Table 2.1. Analysis of the protection capability area in Scenario 1 at time $T = 10$ years. The original values with $\alpha = 1.0$ and $x_0 = 0.7$ are bolded. Columns with $x_0 = 0.58$ and $\alpha = 0.85$ show the two alternative calculations.

Protection, Scenario 1, $T = 10$ years			
α	**1**	1	0.85
x_0	**0.7**	0.58	0.7
X_1	**0.13**	0.13	0.15
X_2	**0.18**	0.18	0.21
X_m	**1.12**	1.00	1.00
X_k	**0.63**	0.58	0.70
A	**0.29**	0.29	0.34
$X = X_m - x_0$	**0.42**	0.42	0.30
$P_{prot} = XA + x_0$	**0.82**	0.70	0.80

2.5. Optimal Number of System Units

Military capability is often defined as the amount of armament. However, different classes of armament are difficult to quantify. The Quantified Judgment Method (QJM) is a mathematical method quantifying weapons, ordnance, fighters etc. (Dupuy 1985). The QJM method is based on historical data and scoring different armaments.

Next, we develop a general modelling framework for optimising the capability of functionalities or systems in a system of systems used in operations or missions (P III). We present a mathematical model for calculating the optimal number of systems to meet capability demand when a particular demand can be addressed by multiple distinct systems. In the model budgeted resource constraints, saturated and non-saturated capabilities and quantity discounts are considered. The model allows for any number of capability areas and systems. A numerical example is provided by the usual case of three capability areas and two systems introduced in (P I).

We generalise our earlier work (P I) to the desired level of granularity of functionalities. We extend the modelling method for multiple systems and an unlimited number of functionalities (P III). In our earlier work, we modelled systems like satellites and UAVs without defining explicitly how many satellites and UAVs are deployed in different scenarios and planning horizons. Now, our main objective is to develop a model for quantifying the optimal number of system units. This is a common problem in many fields of military, economy, business and public organisations. The method can be used to model embedded systems as new systems are joined with an existing system of systems. The existing systems are divided into two sets of functionalities, parallel and serial, with the new system functionalities.

Two different domains of operations are identified: the number of units is either saturated or non-saturated. In the saturated state, the number of units exceeds a threshold value where the number of units is so high that more than one unit is available to be used in the operation for the same functionalities. That is, overlapping or redundant capabilities are available. In a military context, this is a desirable state of affairs. In the model, ratios of the number of units are independent of the total budget of funds. So ratios of the count of system units can be studied without knowing the exact amount of available funds. If the budget is known, the actual number of system units can be calculated from the model. We also show how to incorporate a model for quantity discounts as a mathematical formula or as an empirical data set. (P III)

Our earlier version of the model (P I, P II) was designed only for two systems and their combined use. In (P III), we present general methods of modelling multiple systems composing of several different functionalities. The functionalities can be parallel or serial when viewed on a capability level (see Fig. 2.3). The necessary functionalities are as serial sub-systems and the alternative functionalities are as parallel sub-

systems. The granularity of the model is decided on the grounds of modelling requirements and the fidelity of the input data. An example of three systems consisting of a different number of sub-systems, or functionalities, is shown in Fig. 2.3. The first system has three sub-systems, one of which is common for all three systems (oval symbol), one is common for two systems (triangle symbol) and one is not included in the other two systems (diamond symbol).

In the capabilities portion of Fig. 2.3, the three systems are disassembled and recombined as serial and parallel sub-systems to form the structure of the capabilities. Here, capabilities can be capability areas as in (P I) or lower-level capabilities needed in the operation. The general form for the total capability value P for $\mathcal{I}$ systems and F functionalities is

$$P = \prod_{f=1}^{F} \left((M_f - p_f) A_f + p_f \right), \tag{11}$$

where p_f is the initial value of functionality f and

$$A_f = 1 - \prod_{i=1}^{\mathcal{I}} (1 - X_{if}). \tag{12}$$

System capabilities are expressed as

$$X_{if} = 1 - (1 - x_{if})^{n_i}, \, if \, n_i \geq n_i' \tag{13}$$

and

$$X_{if} = n_i x_{if}, \, if \, n_i < n_i', \tag{14}$$

where x_{if} is the capability value of functionality f in system i. The threshold for system i is denoted by n_i'. Below the threshold value, capabilities are additive, and above the threshold value, the combined use of systems is taken into account according to the probabilistic formula of non-exclusive events. (P III)

The method of the Lagrangian function is used for deriving the optimal number of system units. The method is a general mathematical procedure for solving optimal system states. Theoretically, different models for capabilities or cost functions could be used instead of the choices taken here. In this case, the Lagrangian function has the following form

$$\mathcal{L} = \prod_{f=1}^{F} \left((M_f - p_f) A_f(n_1, \dots, n_{\mathcal{I}}) + p_f \right) + \lambda \left(C - \sum_{i=1}^{\mathcal{I}} g_i(n_i) \right), \tag{15}$$

where λ is the Lagrangian multiplier and $g_i(n_i)$ is the cost of the number of n_i systems of system i. Different forms for the function $g_i(n_i)$ can be used. The linear functional form is often a realistic choice. (P III)

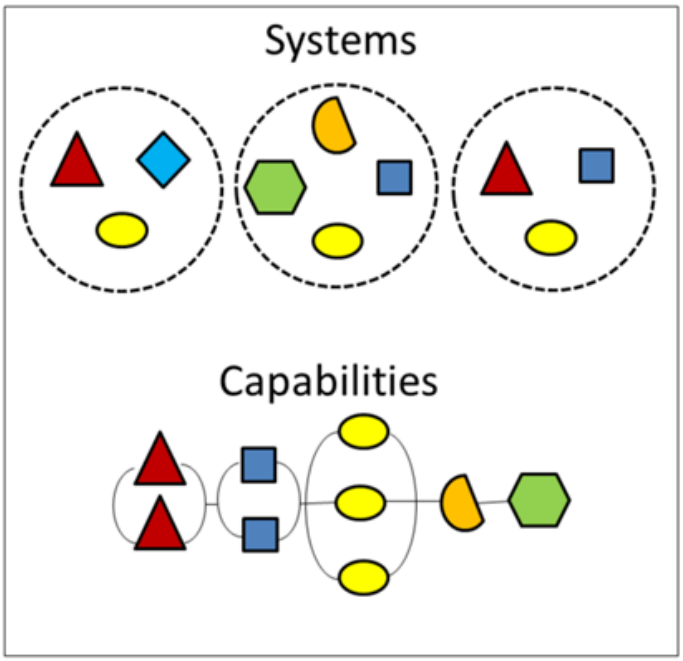

Fig. 2.3. Three systems consisting of six different functionalities (10 sub-systems). One functionality (diamond symbol) has no effect on the capability under investigation.

In (P III), we discuss also how to consider quantity discounts or how to use empirical data based on specific agreements between a vendor and a purchaser. The linear and discount versions of $g_i(n_i), i = 1, \dots, \mathcal{I}$ are

$$g_i^{(lin)}(n_i) = h_i n_i \tag{16}$$

and

$$g_i^{(dis)}(n_i) = h_i n_i^{1-\delta_i}. \tag{17}$$

where δ_i is a tuning parameter for the discount amount and h_i is the cost of one system. The discount version $g_i^{(dis)}(n_i)$ is an example of possible functional forms of discount functions that should be replaced by the function for the system i in the actual procurement situation.

Taking derivatives with respect to λ and n_i by the standard Lagrangian method we have that C is the total cost of all systems and the number of each system $i = 1, \dots \mathcal{I}$ obeys the second formula below

$$C = \sum_{i=1}^{\mathcal{I}} g_i(n_i) \tag{18}$$

and

$$\sum_f (M_f - p_f) \frac{\partial A_f}{\partial n_i} = \lambda g_i'(n_i), \tag{19}$$

where

$$-ln(1-x_{if})(1-x_{if})^{n_i}\prod_{\substack{i=1\\ i\neq n_i}}^{\mathcal{I}}(1-x_{if})^{n_i}=\lambda g_i'(n_i),\text{ if } n_i\geq n_i' \tag{20}$$

and

$$-x_{if}\prod_{\substack{i=1\\ i\neq n_i}}^{\mathcal{I}}(1-n_i x_{if})=\lambda g_i'(n_i),\text{ if } n_i< n_i'. \tag{21}$$

These can be expressed with A_f as

$$-ln(1-x_{if})A_f=\lambda g_i'(n_i),\text{ if } n_i\geq \mathrm{n}_i' \tag{22}$$

and

$$\frac{x_{if}}{1-n_i x_{if}}A_f=\lambda g_i'(n_i),\text{ if } n_i< n_i'. \tag{23}$$

Taking ratios of the previous formulas we obtain (P III)

$$\frac{g_{i_1}'(n_{i_1})n_{i_1}}{g_{i_2}'(n_{i_2})n_{i_2}}=\frac{\sum_{f=1,F}(M_f-P_f)\,(1-A_f)Z_{i_1,f}}{\sum_{f=1,F}(M_f-P_f)\,(1-A_f)Z_{i_2,f}},\, i_1,i_2=1,\dots\mathcal{I}, \tag{24}$$

where we have denoted the saturated and non-saturated alternatives by

$$Z_{if}=-ln(1-X_{if}),\text{ if } n_i\geq n_i' \tag{25}$$

and

$$Z_{if}=\frac{X_{if}}{1-X_{if}},\text{ if } n_i< n_i'. \tag{26}$$

It is noteworthy that systems can be in different states of saturation in one scenario, that is, n_i' is different for each system $i=1,\dots\mathcal{I}$. We denote

$$R_i=\sum_{f=1,2,3}(M_f-P_f)(1-A_f)Z_{i,f},\, i=1,\mathcal{I}. \tag{27}$$

In the case of linear price function, we have the ratio

$$\frac{n_{i_1}}{n_{i_2}}=\frac{h_{i_2}}{h_{i_1}}\frac{R_{i_1}}{R_{i_2}}. \tag{28}$$

We illustrate the model with three functionalities $f=1,2,3$ and two systems $i=1,2$ in (P III).

3.

COMBAT MODELLING

We present a novel combat equation for calculating the probabilities of victory of individual battles between two opposing sides. Durations of battles are predicted with a related model of decision boundaries used as describing stopping rules of a battle. The proposed models are useful in explaining and understanding combat phenomena together with other mathematical and non-mathematical methods.

3.1. Probability of Victory

Operational insights and predictions of the battle outcome can be achieved by using models in analysing combat situations (Low 1995; Bracken 1995; Hartley 2001; Hartley 1995). We discuss macroscopic models that describe battles on a high functional level. Typical variables of these models are the force sizes of attackers and defenders at the beginning and the end of battles. Force sizes govern the situation and they also reflect other variables of the battle. We present a new model (P IV) based on a combination of force strengths and breakpoints, which are implied by a win/lose decisions methodology of two opposing sides. The model is applied to empirical data from a large ensemble of different battles, and for the evolution of three historical single battles of Kursk, Ardennes and Incheon-Seoul. In the analysis of these three battles, variances of the force sizes are calculated from the empirical data, and variance is not a model parameter.

One novelty of our model (P IV) is in predicting the probability of victory in a particular case of an attacker and a defender. A macroscopic combat equation is derived with stochastic methods for log-normally distributed force sizes. The model has two parts, the model for stochastic processes and the model for decision boundaries. Geometric Brownian motion and linear time-dependent functions are used in the modelling. Decision boundaries have a complex dependency on situational conditions and decision rules. To demonstrate the method with analytical mathematical formulas, a simplified linear model has been used for the decision boundaries.

Parameters of the model are variances of the attrition processes, slopes of the decision boundaries and advantage measures for the attacker and the defender. Model parameters and decision boundaries have real-world interpretations (Jaiswal & Nagabhushana 1995). Typical parameter values for the slopes of decision boundaries and the advantage measures are presented. The model gives the probability to win a battle for individual battles. In this respect, the model is more detailed than earlier phenomenological models.

In the model (P IV), the battle ends when the force size of the losing side declines below the lower decision boundary. At this point, the losing side surrenders, or the battle ends in some other way. At the same time, the force size of the winning side is assumed to reach the upper boundary value. The decision boundaries are also called stopping boundaries. The losing side and the winning side may exchange the roles of losing and winning during combat. In the course of combat, we can call the decision boundaries upper and lower boundaries because the winner is not known before the end of the combat.

A simple model for decision boundaries is presented with linear time dependence (P IV). Human factors affect the decision process, which makes the modelling more difficult. No empirical data are available, but it might be possible to get expert opinions about the decision parameters and their values. Another way to look at the question is to construct a model for the decision boundaries and to try to make conclusions about the model parameters consistent with the available empirical attrition data.

One successful macroscopic mathematical expression describing the historical data is the Helmbold relationship (Helmbold 1987). Dean S. Hartley III has fit the parameters of the Helmbold relationship from empirical data of about 800 battles (Hartley 2001). The model applies to ancient, and to some degree, to modern battles. No theory or modelling is explaining the formula. The Helmbold relationship is

$$ln\left(\frac{x_0^2 - x^2}{y_0^2 - y^2}\right) = \alpha ln\left(\frac{x_0}{y_0}\right) + \beta, \tag{29}$$

where $x = x(t)$ and $y = y(t)$ are force sizes of the opposing forces at time t and $x_0 = x(0)$ and $y_0 = y(0)$. On the right side of the equation, α and β are constants. One of the main differences in comparison with the proposed model is that the Helmbold relationship is for the average reference line that separates the empirical data points into two sets. The new formula is for the individual data points. Since results from the model of this study are probabilities while the other is not, the two models cannot be directly compared. Both models predict the winner of combat in about 79 per cent of the cases (Hartley 2001).

In reality, the decision boundary for the losing side to surrender varies during the battle. The simplest model is a linear time-dependent deterministic function. When no reinforcement occurs, the force sizes of the two sides are descending, and this implies descending stopping boundaries. Usually, the two sides don't immediately use all the available force against the enemy. In the model, the effective force sizes may be smaller than the documented force sizes of the historical combat data. If effective force sizes are used, stopping boundaries may be constant, or even ascending, or the slopes of the descending boundaries may be smaller. The model presented allows both types of data, real or effective force size data. Also, aggregated data of manpower and weaponry can be used (Bracken 1995; Turkes 2000).

We derive the combat equations by considering stochastic processes where the force sizes of fighting troops are modelled as geometric Brownian motion. Stochastic models don't usually give closed-form formulas for probabilities or force sizes. In the particular case of geometric Brownian motion, the results can be presented in closed-form equations. No numeric calculations or simulation are needed in the analysis. Lanchester equations (Lanchester 1914; Osipov 1915) have stochastic equivalents usually based on discrete-time Markov chains (Morse & Kimball 1951). In practical combat situations, numerical computations of stochastic Lanchester equations are needed because no simple closed-form formula exists.

The proposed combat equation is a function of the initial force sizes, attrition rates, variances, and decision boundaries of two opposing forces. If the attacker or the defender has an additional advantage in the battle, which is not included in other features of the model, a parameter adjusting the decision boundaries is available. This parameter is also useful in asymmetrical battles.

The combat model consists of two parts. Firstly, the attrition process is modelled as a stochastic process. Several alternatives exist - Brownian motion, geometric Brownian process, Poisson jump process, and so on (Karlin & Taylor 1975; Florescu 2015). Secondly, a model for the decision boundaries is needed. Decision boundaries can be modelled as deterministic or stochastic functions. In this model, we use geometric Brownian motion as the model for an attrition process and a linear function of time as the model for decision boundaries. The linear decision boundary is a natural choice because the empirical data (Hartley 2001) do not enable studying more general time-dependent functions.

We assume that force sizes obey the law of geometric Brownian motion X_t as a function of time t. Stochastic process X_t follows the stochastic differential equation

$$dX_t = X_t dB_t, \tag{30}$$

where Brownian motion B_t follows stochastic differential equation $dB_t = \hat{\mu}dt + \sigma dW_t$. Constant drift is denoted by $\hat{\mu}$ and constant standard deviation by σ (Parameters of the model are listed in Table 3.1). The process W_t as a function of time t is a standard Brownian motion normally distributed with mean zero and variance t denoted by $W_t \sim N(0, t)$. Stochastic changes in X_t are proportional to the values of the quantity itself. Owing to this property, geometric Brownian motion has been commonly used to model stock values. This can be justified in combat modelling also because attrition processes have similar properties.

Table 3.1. List of model parameters for attacker and defender sides.

$\boldsymbol{\sigma_A^2}$ and $\boldsymbol{\sigma_D^2}$	Variances of the logarithmic attrition data
$\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$	Parameters determine the slopes of decision boundaries
$\mathfrak{p}_A$ and $\mathfrak{p}_D$	Advantage parameters of decision boundaries

Geometric Brownian motion is one of the few stochastic processes that can be solved in closed-form formulas. The stochastic differential equation for geometric Brownian motion X_t can be solved using the Itô formula (Karlin & Taylor 1975; Florescu 2015) giving

$$X_t = x_0 exp\left[\left(\hat{\mu} - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right]. \tag{31}$$

Later, we denote the drift term with the Itô effect in parenthesis as $\mu = \hat{\mu} - \frac{1}{2}\sigma^2$. Now, we have the model for both sides of combat, the attacker's force size is X_t and the defender's force size is Y_t. We denote the force sizes at the end of the combat at time $t = T$ by $X_T = x_u$ for the attacker and by $Y_T = y_d$ for the defender. The corresponding initial values at time $t = 0$ are denoted by x_0 and y_0.

Usually, time dependencies of attrition processes are not known and most of the empirical data are for a one-day duration. This is why the variances of the attrition processes cannot be calculated from the data and we have to take the variation of processes as a free parameter. We take the duration of a battle as our time unit and set time $T = 1$. This means that attrition rates and variances are calculated for the duration of the combat. In the case of geometric Brownian motion, the variances are calculated from the logarithmic data and are commonly called volatilities. We denote

$$\mu_A = ln\left(\frac{x_u}{x_0}\right) \tag{32}$$

and

$$\mu_D = ln\left(\frac{y_d}{y_0}\right). \tag{33}$$

We propose the following model for the slopes of decision boundaries for the attacker and the defender correspondingly:

$$M_A = -(1-\alpha)\mu_D - \alpha\mu_A \tag{34}$$

and

$$M_D = -(1-\beta)\mu_A - \beta\mu_D. \tag{35}$$

By inserting μ_A and μ_B we have

$$M_A = -\alpha ln\left(\frac{x_u y_0}{x_0 y_d}\right) + ln\left(\frac{y_0}{y_d}\right) \tag{36}$$

and

$$M_D = -\beta ln\left(\frac{x_u y_0}{x_0 y_d}\right) + ln\left(\frac{x_0}{x_u}\right). \tag{37}$$

In the derivation of the combat equation, the martingale property of geometric Brownian motion and the optional stopping theorem of stochastic analysis is needed (Karlin & Taylor 1975). The formula for the probability of the attacker to win P_A is (P IV)

$$P_A = \frac{x_0^{-|m_A|} - e^{-|m_A|M_A} x_u^{-|m_A|}}{e^{-|m_A|M_A} x_u^{-|m_A|} - e^{-|m_A|M_A} x_d^{-|m_A|}}. \tag{38}$$

The main result in (P IV) for the probability P_A for the attacker to win can be expressed as

$$P_A = \frac{(S^{\alpha-1})^{-|m_A|} - A^{-|m_A|}}{1 - A^{-|m_A|}}, \tag{39}$$

where

$$S = \frac{x_u y_0}{x_0 y_d}, \tag{40}$$

$$m_A = \frac{-2}{\sigma_A^2}(\mu_A - M_A) = \frac{-2}{\sigma_A^2}(1-\alpha) ln(S) \tag{41}$$

and

$$A = S^{\alpha-1}(1 + \mathfrak{p}_A) - \mathfrak{p}_A. \tag{42}$$

Parameter $\mathfrak{p}_A$ determines the initial position of the lower boundary (see Fig. 2 in (P IV)) and describes the initial advantage or disadvantage of the attacker. Values greater than one stand for advantage and values less than one for disadvantage. We have the corresponding equation for the defender to lose as

$$P_D = \frac{\left(S^{1-\beta}\right)^{-|m_D|} - D^{-|m_D|}}{1 - D^{-|m_D|}}, \tag{43}$$

where

$$m_D = \frac{-2}{\sigma_D^2}(\beta - 1) ln(S) \tag{44}$$

and

$$D = S^{1-\beta}(1 + \mathfrak{p}_D) - \mathfrak{p}_D. \tag{45}$$

Parameter $\mathfrak{p}_D$ for the defender corresponds to $\mathfrak{p}_A$ for the attacker. The probabilities of the attacker winning and the defender losing should be equal. This means that the parameters of the attacker and defender sides are related by the equation $P_A = P_D$. In (P IV) these are called the attacker view and the defender view, but both views give the same results because the parameters are synchronised. When S approaches one

$$\lim_{S \to 1} P_A = \frac{\mathfrak{p}_A}{1 + \mathfrak{p}_A} \tag{46}$$

and

$$\lim_{S \to 1} P_D = \frac{\mathfrak{p}_D}{1 + \mathfrak{p}_D}. \tag{47}$$

Note, that formulas for m_A and m_D can be expressed as functions $\hat{\mu}_A$ and $\hat{\mu}_D$ or as functions μ_A and μ_D as

$$m_A = 1 + \frac{-2}{\sigma_A^2}(\hat{\mu}_A + M_A) = \frac{-2}{\sigma_A^2}(\mu_A + M_A) \tag{48}$$

and

$$m_D = 1 + \frac{-2}{\sigma_D^2}(\hat{\mu}_D + M_D) = \frac{-2}{\sigma_D^2}(\mu_D + M_D). \tag{49}$$

The difference of one is caused by the Itô calculus of stochastic processes (Karlin & Taylor 1975; Florescu 2015). In Fig. 3.1, we investigate the impact of this difference for three values of variance σ^2. We assume that attacker and defender attrition processes have the same parameter values, for example, $\sigma^2 = \sigma_A^2 = \sigma_D^2$ and $\mathfrak{p}_A = \mathfrak{p}_D = 1.1$. That is, parameter values in the formulas for P_A and P_D have not been synchronised. The probability for an attacker to win P_A is shown by black curves and the corresponding probabilities for a defender to win P_D are shown by blue curves. To demonstrate the impact of Itô calculus we replace $\hat{\mu}_A$ by μ_A and $\hat{\mu}_D$ by μ_D in the formulas (that is, one added to m_A and m_D). We see that for high variance values $\sigma^2 = 0.025$ there is a difference of $2\ \%$ or less when S is between 0.9 and 1.1.

Note that the curves illustrating the magnitude of the Itô effect are shifted to opposite directions from the attacker view (A in Fig. 3.1) and the defender view (D in Fig. 3.1). At lower variance values the terms having variance in the dominator dominates and the Itô effect is not significant. On the other hand, because P_A and P_D should be equal, we can see from Fig. 3.1 that after synchronising the two formulas for individual data points the effect of the Itô calculus is cancelled. We conclude that, on average, Itô effect is not significant, but for individual combats with high variance, the Itô calculus can affect the parameter values used for the two opposing sides in the model.

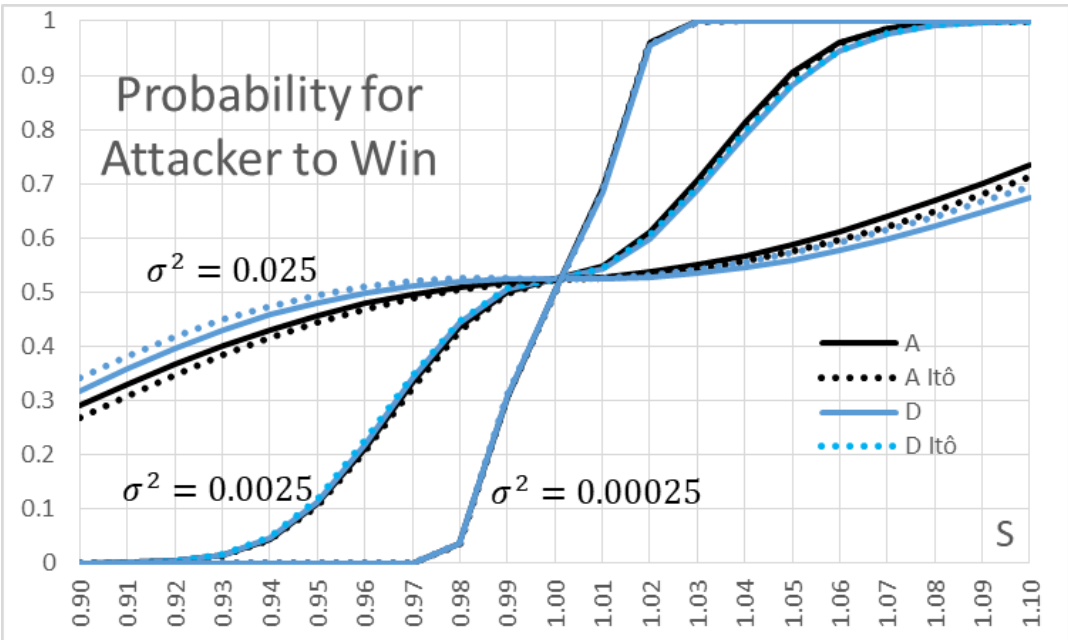

Fig. 3.1. The impact of Itô calculus and the difference between attacker and defender views without synchronisation is illustrated for three variance values as a function of the ratio S. Solid lines show the results from the theory and dotted curves show the effect of Itô calculus. Attacker side results P_A are shown by black curves and defender side results P_D are shown by blue curves.

The model gives the probability of victory of a specified battle between an attacker and a defender. A wide range of variance values in the theory agree with the empirical probability formula. Good compliance of the theory and the empirical data with $\sigma^2 = 0.00025, 0.0025$ and 0.025 can be achieved. The number of correctly predicted victors is 438, or more, out of 593 combats in the dataset. The use of the parameter value of $\mathfrak{p}_A = 1.1$ describing an advantage for attackers provides only a minor improvement to the results compared to the value of no advantage $\mathfrak{p}_A = 1.0$ (see Table 3.2).

Table 3.2. Typical model parameter values of the model with correctly predicted numbers of victors (#).

σ^2	p_A	#
0.00025	1	438
0.0025	1	438
0.025	1	438
0.00025	1.1	441
0.0025	1.1	441
0.025	1.1	432

As the empirical data set has only the initial and final values of force sizes, variances cannot be calculated from the data. However, time-dependent empirical data from three historical combats are available from Ardennes (Bracken 1995), Kursk (Turkes 2000) and Incheon-Seoul (Hartley & Helmbold 1995) campaigns or battles. Variances for the duration of these battles are shown in Table 3.3 (Ardennes battle with manpower and generalised combat power data). Variance values are calculated as variances of logarithmic attrition data. Table 3.3 shows variance values for the attacker and the defender. Variance values are high when compared to the typical values of Table 3.2. Variances for one day can be easily obtained by dividing the variance values in Table 3.3 by the duration of a battle. Attrition rates and volatilities must be calculated using the same time units.

Table 3.3. Variance values of historical attrition data for the duration of battles. Duration is shown in the last column. Combat power includes manpower and weaponry with weighting factors. T4, T5 and T15 refer to (Tables 4 and 5 in Appendix in (P IV)).

Battle	**Attacker Variance**	**Defender Variance**	**Duration In days**
Kursk (T15)	0.0029	0.098	15
Ardennes manpower (T4)	0.335	0.096	33
Ardennes combat power (T5)	0.400	0.156	33
Incheon-Seoul	0.024	0.160	19

Combat equations are symmetric for the attacker and the defender. From the decision boundary model, it follows that the probability of winning combat is a function of ratio S. Quantity S is a good indicator of the victor of a battle. When < 1, the defender is predicted to win the battle, and if $S > 1$ the attacker is predicted to win the battle. The other models mentioned earlier, the Helmbold relationship (Helmbold 1987), the Willard relationship (Willard 1962) or the empirical formula (Hartley 2001) give a similar number of correct predictions for the victor. All three models, and the model in this study, predict the same 438 records correctly and 114 incorrectly. Even 41 cases are not included in this comparison.

In some cases, the value of S is very near or exactly 0.5 and some parameter values of the model give 438 – 441 correct predictions. This cannot be used to calibrate the model parameters, because of the variations of the empirical data and combat situations. Consequently, it is easy to give an empirical formula, without the probability value of victory, predicting the winner in about 438 of 552 cases.

The Ardennes Campaign, commonly known as the Battle of the Bulge, caught the Allies by an almost complete surprise. In the following, Blue denotes the Allies (the United States and Great Britain) and Red denotes the Germans (Fig. 3.2). However, Red attacks suffered from several major weaknesses: personnel quality was weak, transportation networks and air forces were inadequate. After several days of Red penetrations, the Blue forces rallied to slow down and then to stop the Red attacks. By the beginning of day 4, the Blue forces attacked into the left shoulder of the bulge to relieve beleaguered Blue units. By day 11, the Blue air supremacy was brought to bear on the Red units in the bulge. The Blue forces counterattacked, and two weeks later they restored the front line in the Ardennes. Comparing the Ardennes Campaign description and Fig. 3.2, the main operations can be connected with the probability changes. The actual casualty and reinforcement numbers are used in the analysis with the textual information. (Turkes 2000)

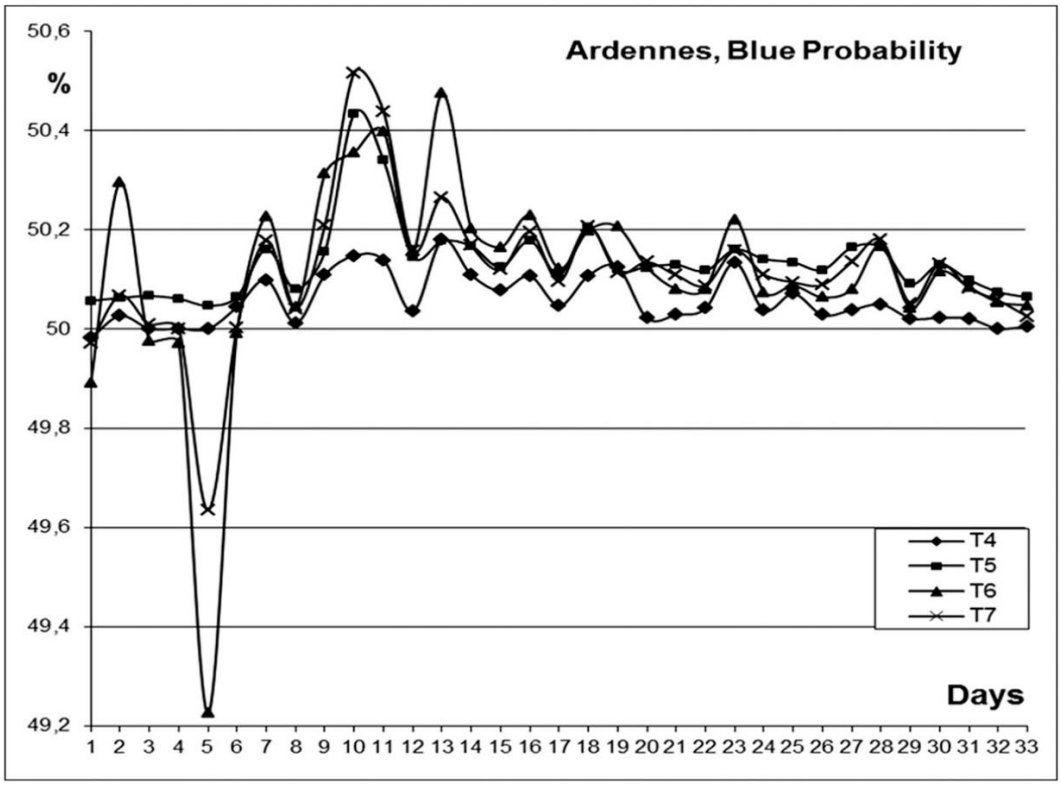

Fig. 3.2. Analysis of the battle of Ardennes. Probabilities for the Blue forces to win the battle during the 33 days of the battle. Every day is handled as an individual combat. Parameter value $\mathfrak{p}_A = 1.0$ has been used.

3.2. Duration of Combats

In (P V), we have discussed a combat model based on geometric Brownian motion with moving decision boundaries. Attrition processes were modelled as actual force sizes separately for the two opposing forces. Decision boundaries were modelled as linear functions of time. The slope for the attacker's decision boundary was determined by the slope of the defender's force size and vice versa. This model has the property of descending decision boundaries when opposing sides are losing manpower. In (P V), we show how the modelling in (P IV) can be carried out with different decision boundaries. Also, our goal is to extend the model to predict combat durations and to compare predicted durations with the observed empirical durations of battles.

We compare using the same methodology, on an aggregate level, theoretical results of probabilities to win battles and expected durations with empirical data. Modelling is done at the detailed level of individual battles but results are studied on the aggregate level. We assume that the empirical data obeys roughly a general pattern and so characteristic model parameter values can be used to describe the entire dataset. The empirical data does not contain enough information to determine model parameter values for individual battles

We define a stochastic process as the difference between the attacker's attrition process and the defender's attrition process. The drift of this process is defined correspondingly as the difference between attacker's drift and defenders drift. A new model for decision boundaries is defined that enables us to predict durations of battles (P V). In the new model, decision boundaries are considered to be constants as a function of time. In a way, the new model of decision boundaries and our earlier work can be considered as two manners of approaching the same problem. The definition of a decision boundary is related to the definitions of attrition processes and drift terms.

The stochastic process describing the combined attrition process of a battle is approximated with a geometric Brownian motion. The new modelling enables predicting both probabilities and durations of battles. Comparing both of these results, probabilities and durations, with empirical data, can reveal new properties of combat effects. As attrition processes are modelled with geometric Brownian motions, the drift parameter has the standard mathematical definition provided in the literature (Karlin & Taylor 1975; Florescu 2015).

In the article (P V), we illustrate the effects of different parameter values with a simplified model (Model 1 in (P V)) where decision boundaries are δx_o above or below the initial force size value. The value of constant δ is approximately 10 % in the empirical data. Three general observations about expected durations of battles are made from the analytical formulas of Model 1 concerning different variances, different decision boundaries and asymmetrical decision boundaries (see Fig. 3.4).

Fig. 3.3A shows that near $S = 1$ values of expected duration $E(T)$ are higher for low values of variance σ^2. This is a direct consequence of the functional form of Equation (9) in (P V). This effect prevails only when σ^2 is approximately between 0.8 and 1.25. Outside this interval the expected duration $E(T)$ is almost independent of the variance value. The interpretation is clear: higher variance means heavier fighting and a shorter duration of the battle.

In Fig. 3.3B a similar comparison with wider decision boundaries is shown. The variance value of $\sigma^2 = 0.01$ is used and the effects of decision boundaries symmetrically 10 %, 15 % and 20 % above and below the initial value of force size x_0 are shown. With wider decision boundaries $E(T)$ is higher. This occurs with all values of S unlike in the curves shown in Fig. 3.3A. The interpretation is that the adversary is more resilient and is not ready to surrender. The battle takes a long time and a higher expected value of duration follows.

In Fig. 3.3C an interesting phenomenon can be seen when decision boundaries are not symmetrical. In this case, the upper decision boundary is higher than the corresponding symmetrical upper decision boundary concerning x_0 would be. In Fig. 3.3C $\delta_{xu} = 0.1$ is the symmetrical case and $\delta_{xu} = 0.2, 0.3$ and 0.4 are asymmetrical cases. The value for the lower boundary $\delta_{xd} = 0.1$ is used in all the cases. The consequence of non-symmetricity is asymmetrical behaviour also in $E(T)$. With high values of S the expected duration value of a battle is higher when compared with the symmetrical case. It is noticeable that effects with low values of S are negligible. The interpretation is that when $S > 1$ the attacker is superior and probably closer to the victory, that is, closer to the upper decision boundary and changes on the upper boundary are more important than changes on the lower boundary far away down. The situation is reversed when $S < 1$.

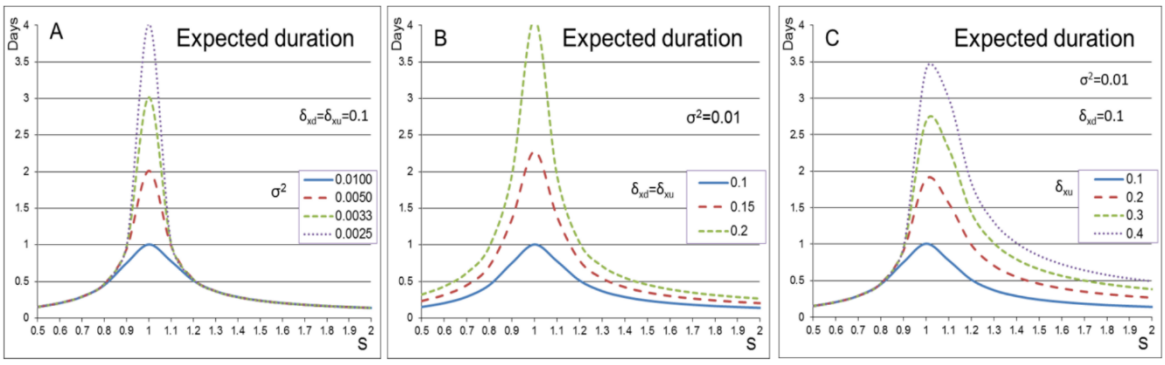

Fig. 3.3. Effects on expected combat durations from Model 1 Equation (9) in (P V). A) Different variances, B) Different decision boundaries, C) Asymmetrical decision boundaries.

Decision boundaries depending on a particular condition of a battle may predict outcomes of battles better than Model 1 in (P V). Consequently, we define (Model 2 in (P V)) upper and lower decision boundaries for the attacker as

$$X_u = x_0 + d_{xu}(x_0 - x_u),\ X_d = x_0 - d_{xd}(x_0 - x_u) \tag{50}$$

and for the defender as

$$Y_u = y_0 + d_{yu}(y_0 - y_d),\ Y_d = y_0 - d_{yd}(y_0 - y_d). \tag{51}$$

Using these decision boundaries the probability to win a battle is (P V)

$$P = \frac{\left(\frac{x_0}{x_0 + d_{xu}(x_0 - x_u)}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}\left(1 - \left(\frac{x_0 - d_{xu}(x_0 - x_u)}{x_0}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}\right)}{1 - \left(\frac{x_0 - d_{xu}(x_0 - x_u)}{x_0 + d_{xu}(x_0 - x_u)}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}}, \tag{52a}$$

when $S \geq 1$ and

$$P = \frac{1 - \left(\frac{y_0 + d_{yu}(y_0 - y_d)}{y_0}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}}{1 - \left(\frac{y + d_{yu}(y_0 - y_d)}{y_0 - d_{yu}(y_0 - y_d)}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}}, \tag{52b}$$

when $S < 1$. The corresponding expected duration is

$$E(T) = \frac{1}{\frac{\sigma^2}{2} - ln(S)} \left(ln\left(\frac{x_0}{x_0 - d_{xd}(x_0 - x_u)}\right) - \frac{1 - \left(\frac{x_0}{x_0 - d_{xd}(x_0 - x_u)}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}}{1 - \left(\frac{x_0 + d_{xu}(x_0 - x_u)}{x_0 - d_{xu}(x_0 - x_u)}\right)^{1-\frac{2\,ln(S)}{\sigma^2}}} ln\left(\frac{x_0 + d_{xu}(x_0 - x_u)}{x_0 - d_{xu}(x_0 - x_u)}\right) \right), \tag{53a}$$

when $S \geq 1$ and

$$E(T) = \frac{1}{\frac{\sigma^2}{2} + ln(S)} \left(ln\left(\frac{y_0}{y_0 - d_{yd}(y_0 - y_d)}\right) - \frac{1 - \left(\frac{y_0}{y_0 - d_{yd}(y_0 - y_d)}\right)^{1+\frac{2\,ln(S)}{\sigma^2}}}{1 - \left(\frac{y_0 + d_{yu}(y_0 - y_d)}{y_0 - d_{yd}(y_0 - y_d)}\right)^{1+\frac{2\,ln(S)}{\sigma^2}}} ln\left(\frac{y_0 + d_{yu}(y_0 - y_d)}{y_0 - d_{yd}(y_0 - y_d)}\right) \right), \tag{53b}$$

when $S < 1$.

Next, we compare the results of the model with the empirical combat data (Hartley 2001). The quality of the empirical data is not high and also the data is scarce for low and high values of $S \notin (0.9,1.1)$. Also, many battles have very long duration or extremely high casualties. Similar durations can be achieved with high variance values or narrow decision boundaries and low variance values or wide decision boundaries. As a consequence, it is not possible to fit the parameters of the model with the empirical data. In the following, we compare the theoretical results, probabilities and expected durations, with the empirical data and try to find typical parameter values that are consistent with the theory.

Fig. 3.4 shows that the variance value of $\sigma^2 = 0.0025$ complies with the empirical data. The value $\sigma^2 = 0.0025$ is in good agreement with Fig. 3 in (P IV) using a different model for attrition processes and decision boundaries. We know that individual

battles can have lower or higher variances, especially higher variances that are more abundant for high or low values of S. Also, we know that some battles with $S \approx 1$ can have variance values higher than $\sigma^2 = 0.0025$, still consistent with the theory. The quality and quantity of the empirical data did not permit analysis of the empirical data categorised in intervals of observed durations.

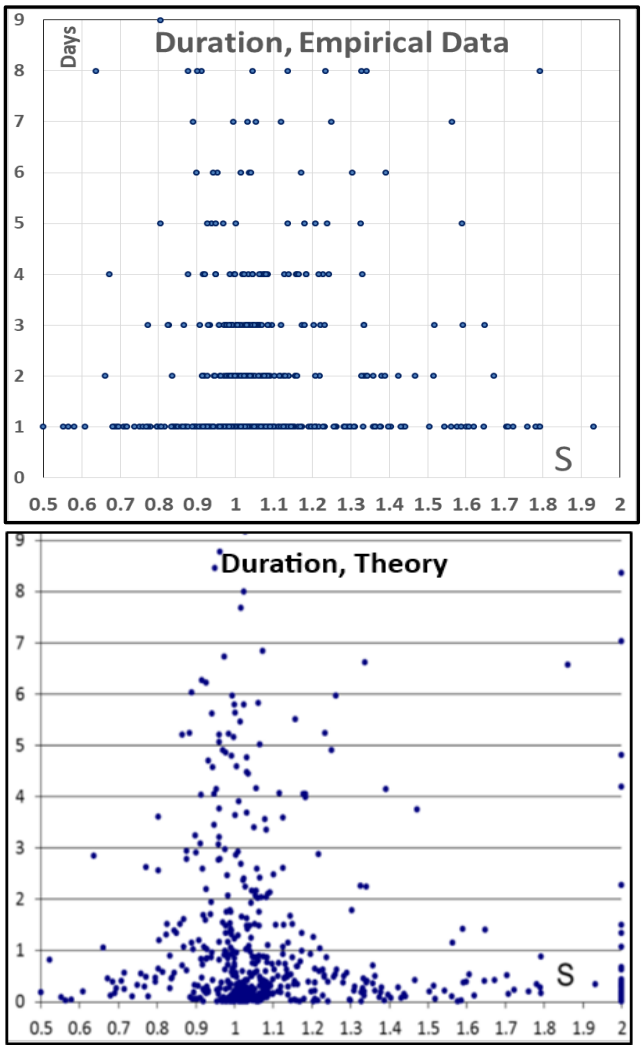

Fig. 3.4. Empirical durations (top) from the data set published in (Hartley 2001) and theoretical expected durations (bottom) with model parameter values $\boldsymbol{\sigma^2 = 0.0025}$ and $\boldsymbol{d_{xy} = d_{xd} = d_{yu} = d_{yd} = 1.0}$.

We have concluded in (P IV) that the model with asymmetrical decision boundaries is in better compliance with the empirical combat data than with symmetrical decision boundaries if attackers are superior to defenders. This suggests that attackers have an additional advantage over defenders. Asymmetrical features between attackers and defenders are discovered by comparing empirical data with the modelling results of probabilities and expected durations. Higher variances alone can explain theoretical probability values of victory but asymmetrical decision boundaries are needed to explain theoretically expected duration values to match the empirical combat durations.

4.

MODELLING NETWORKED SYSTEMS

The application of new information technology in operational systems and communication networks increase situational awareness in network-centric operations. Information sharing, information access and speed can provide an advantage to a party in modern conflict. In this study, we present models for describing the utility of networked information services, the resilience of communication networks and optimal placement of sensors in a network structure.

4.1. Metrics for Networked Systems

Modelling networked systems of services and users in a military context provide vital information for the defender in planning more robust networks against targeted attacks. Mathematical methods of this study (P VI) use analytical results of network theory and they are not limited to low failure probabilities. This property is important in military applications and also in unexpected rare events.

Much of the modelling is based on connectivity in a network structure. One connection is enough for two parties to communicate, two connections are enough for three parties and so on. Maintaining a connection means that there is a path composed of one or more links between nodes. In case the nodes or corresponding links connecting the nodes are not functioning during the same period, the system is down and it is not in use. For the most part, our examples use nodes but similar models can be presented for links. Both node and link failures can occur simultaneously in real situations and it is possible to model both of these types of events in the same model.

In the following, the networked system is composed of three different types of elements: network links, functional network nodes and technical network nodes. Functional nodes are computer centres, command centres, users or networked services. Technical nodes such as routers and repeaters have no applications available for end-users. The network metrics use the assumption that the functional nodes can perform the same tasks. This is a simplification because we aim to present the modelling methods with mathematical formulas that are easier to write down and understand. The approach taking into account the exact network structure is different from network metrics definitions that sum all possible connections and nodes and weight them with different numerical link and node weight values (Ling 2005).

First, we study a system of nodes and assume that the network is functional. This kind of scenario is realistic, for example, when several servers can perform the same tasks and one or more servers are sufficient to serve users of operative systems. We

present two models, the first is a simple model to illustrate the ideas and the second model is a more general one. In the first model, we assume that the number of servers n is reserved but only one server is sufficient to fulfil the needs of users. Parameters p_z, p_t and p_r describe probabilities of failure, replacement and repair in a time unit, for example in one day. We assume that the process follows a Markov process (Karlin & Taylor 1975) and the following Markov matrix (or stochastic matrix) $\mathbb{M}$ for the system of n nodes (servers)

$$\mathbb{M} = \begin{pmatrix} 1-(1-p_z)p_r-(1-p_z^{n-1})p_t & (1-p_z)p_r & (1-p_z^{n-1})p_t \\ p_z^n & 1-p_z^n & 0 \\ p_z^n & 0 & 1-p_z^n \end{pmatrix}. \tag{54}$$

The rows and columns in the matrix correspond to the number of functioning nodes as no functioning nodes (row/column 1) or at least one functioning node (rows/columns 2 and 3). Rows and columns correspond to states 0, 1 and 2 in the theory of Markov processes. The only difference between states 1 and 2 is that they have a different history.

The limiting distribution (Karlin & Taylor 1975) can be solved from the following steady-state equation

$$\pi_j = \sum_{k=0}^{n} \pi_k \, M_{k,j}, \; j = 0, \dots n. \tag{55}$$

We have for the non-functioning system π_0 and functioning system $\pi = 1 - \pi_0 = \pi_1 + \pi_2$

$$\pi_0 = \frac{1}{1+Q(n)} \; and \pi = \frac{Q(n)}{1+Q(n)}, \tag{56}$$

where $Q(n)$ is the ratio of uptime to downtime of the system:

$$Q(n) = \frac{\pi}{\pi_0} = \frac{(1-p_z)p_r + (1-p_z^{n-1})p_t}{p_z^n}. \tag{57}$$

A useful formula for n can be solved as a function of $Q(n)$ as

$$n = \frac{ln\left(\frac{p_r + p_t - p_z p_r}{p_z Q + p_t}\right)}{ln(p_z)} + 1. \tag{58}$$

In the second model, the number of functioning nodes and the repair time of nodes are considered. We write the Markov matrix for three nodes $\mathbb{M}$ as

$$\begin{pmatrix} (1-p_r)^3 & 3p_r(1-p_r)^2 & 3p_r^2(1-p_r) & p_r^3 \\ (1-p_r)^2 p_z & (1-p_r)^2(1-p_z) + 2p_r(1-p_r)p_z & 2(1-p_r)p_r(1-p_z) + p_r^2 p_z & p_r^2(1-p_z) \\ (1-p_r)p_z^2 & 2(1-p_r)p_z(1-p_z) + p_r p_z^2 & (1-p_r)(1-p_z)^2 + 2p_r p_z(1-p_z) & p_r(1-p_z)^2 \\ p_z^3 & 3p_z^2(1-p_z) & 3p_z(1-p_z)^2 & (1-p_z)^3 \end{pmatrix}.$$

The limiting distribution in the general case (P VI) for n nodes is

$$\pi_i = \binom{n}{i} \frac{p_r^i p_z^{n-i}}{(p_r + p_z)^n}, \; i = 0, \dots, n. \tag{59}$$

A summary of the two models is provided in Table 4.1.

Table 4.1. Summary of Model 1 and Model 2

Model 1	Eq.	Model 2	Eq.
$\pi = \dfrac{p_r + p_t}{\dfrac{p_z^{n-1}}{1 - p_z^{n-1}} + p_{r+}p_t}$	(56)	$\pi = 1 - \dfrac{p_z^n}{(p_r + p_z)^n}$	(59)
$Q(n) = \dfrac{(1 - p_z)p_r + (1 - p_z^{n-1})p_t}{p^n}$	(57)	$Q(n) = \left(\dfrac{p_r}{p_z} + 1\right)^n - 1$	
$n = \dfrac{ln\left(\dfrac{p_r + p_t - p_z p_r}{p_z Q + p_t}\right)}{ln(p_z)} + 1$	(58)	$n = \dfrac{ln(Q + 1)}{ln(p_r + p_z) - ln(p_z)}$	

Fig. 4.1 shows results for both models with the parameter values of $p_r = p_t = 0.1$ and $p_r = 0.4, p_t = 0.0$ for Model 1 in (P VI) and $p_r = 0.4$ for Model 2 in (P VI). The ratio of uptime to downtime Q as the function of node failure probability p_z is shown for $n = 1, \dots, 6$ node systems. Due to the different Markov matrices of the systems, different results from the two models are expected. Markov matrices determine the repair and replacement processes of the models. Model 1 and Model 2 are examples of how to design the model with a Markov matrix with different states of the system and probabilities of state transitions between the states.

The first step is to construct a Markov matrix that describes the real process as closely as possible. Repair and replacement and their combined use can be modelled. In a realistic scenario, it is possible that parameter values p_r and p_t are functions of the number of nodes under repair. Repair workers may be a finite resource causing the probability p_r to decrease as the number of units under repair increase. A more general model would consider a simultaneous repair and replacement process of different equipment with specialised personnel for different types of systems. That kind of model could be used to optimise training and employment of technicians with different combinations of skills because it is not feasible that everybody in the repair team knows how to repair all technical systems of operations.

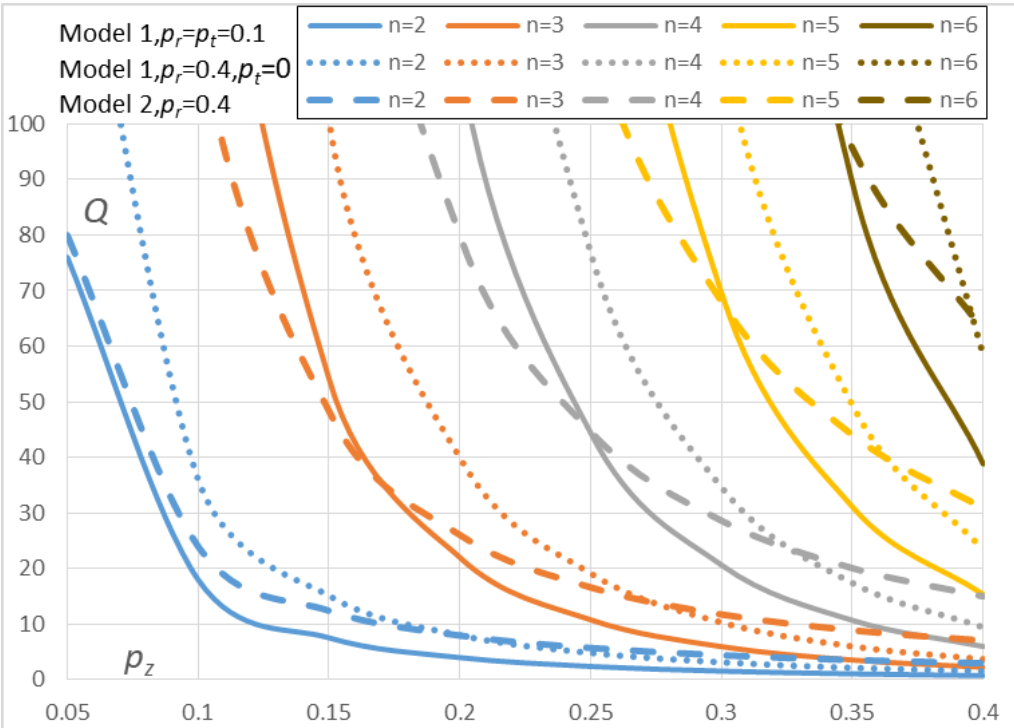

Fig. 4.1. The ratio of uptime to downtime Q as a function of node failure probability p_z for Models 1 and 2. Model parameters are $p_r = p_t = 0.1$ and $p_r = 0.4$, $p_t = 0.0$ for Model 1 and $p_r = 0.4$ for Model 2.

The repair and replacement process should meet requirements for the ratio Q of uptime to downtime. We give one numerical example for the performance requirement of Q. If we assume that Model 1 describes the repair and replacement process, and we have the requirement for $Q = 50$, or less than 2 % of the time not functioning. With the value of failure probability $p_z = 0.2$ and the process characteristic values of $p_r = p_t = 0.1$, we get from the equation of the previous table (bottom left) for the number of nodes $n \approx 3.5$. This means that four nodes are needed to maintain the system at the level of more than 98 % of the time in operation. From Fig. 4.1. we can see that in this case four nodes are needed also for Model 1 with $p_r = 0.4$, $p_t = 0.0$ and $p_r = 0.4$ for Model 2. We see from Fig. 4.1 that three nodes in Model 1 with $p_r = 0.4$, $p_t = 0.0$ are needed for $Q = 40$, or less than 2.5 % of the time not functioning.

One difference between Models 1 and 2 in (P VI) is that it is possible to attain somewhat higher Q values in Model 2 with the same number of nodes (see Fig. 4.1, Model 1 with $p_r = 0.4$, $p_t = 0.0$ and Model 2 with $p_r = 0.4$). The reason is a more efficient repair process of Model 2 compared to that of Model 1.

Next, we study (P VI) the network structure and assume that nodes are functional. Here, we define connectivity as the probability of connection between two nodes in a network. Between neighbouring nodes, link connectivity means the probability of direct connection between the two nodes, that is the probability of a functioning link.

We demonstrate the method of calculating connectivity between network nodes with the example network of Fig. 4.2. The network has four nodes and five links between the nodes. There are three different conceivable node pairs 1-2, 2-4 and 1-3. All other alternatives are similar. For example, connectivity between nodes 1-4, 2-3 and 3-4 are similar to connectivity between nodes 1-2. We assume that the connectivity of links does not depend on their direction.

We assume that all the links in the network are functional with the same probability of p. For the connection between nodes 1 and 2, by using the number of alternative configurations 1, 5, 9, 5 and 1 of Fig. 4.2, we can write the probability of functioning connection as

$$p_{1-2} = p(1-p)^4 + 5p^2(1-p)^3 + 9p^3(1-p)^2 + 5p^4(1-p)^1 + p^5. \tag{60}$$

This can be simplified as

$$p_{1-2} = p + p^2 - 2p^4 + p^5. \tag{61}$$

The corresponding equations for connections between nodes 2 – 4 and 1 – 3 are

$$p_{2-4} = p + 2p^2 - 2p^3 - p^4 + p^5 \tag{62}$$

$$p_{1-3} = 2p^2 + 2p^3 - 5p^4 + 2p^5. \tag{63}$$

As can be seen, connectivity between two nodes in a network is a polynomial of at most the number of links. Links in real networks can have different connectivity depending on direction. This can be dealt with using the directional values of link connectivity. A consequence of this is that the connectivity between nodes 1-2 can be different from the connectivity between nodes 2-1.

Effects of targeted attacks against nodes and links can be investigated with the models of (P VI). We have demonstrated (P VI) that, if locations of services are not known, to cause maximum damage, attacks with high link failure probabilities should be targeted equally on all links of the network. For lower failure probabilities below a threshold value, it is optimal for an attacker to focus attacks against lower degree nodes. Usually, the optimal set of links goes through several intermediate thresholds depending on the network structure. The threshold values of link failure probabilities for different combinations of links of the network are valuable information in planning and constructing robust network structures

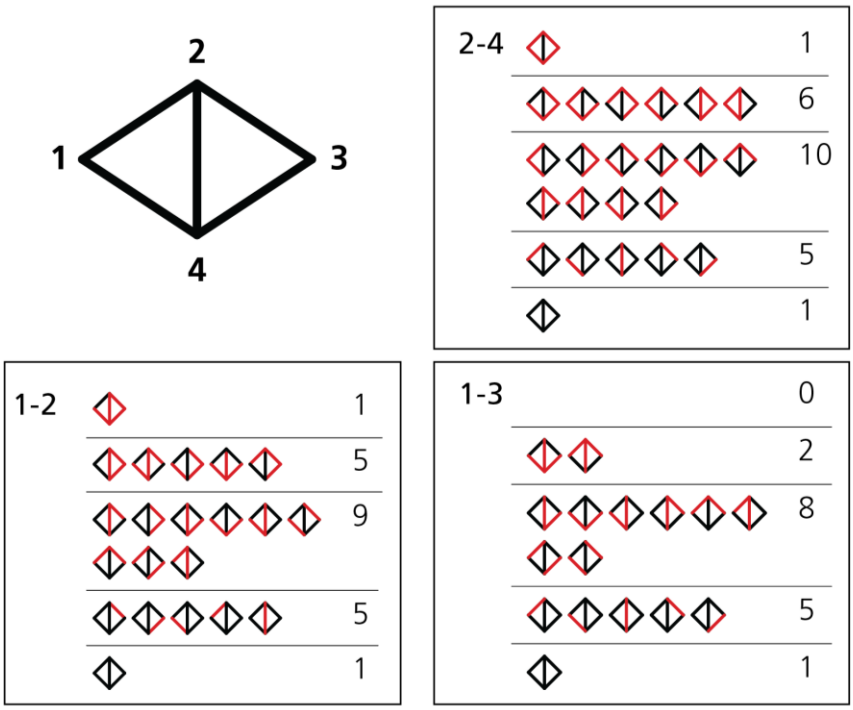

Fig. 4.2. Example network topology of four nodes and five links. Different configurations of connectivity and the number of alternatives between nodes 1-2, 2-4 and 1-3 are listed after the graph.

4.2. Network Resilience and Utility of Services

Modelling networked services in the military context is an important research question. We present a methodology (Kuikka 2019b) for calculating network resilience and service usability measures based on classical reliability polynomials and a method for aggregating layered service structures. Typical case studies are discussed and illustrated with a small real-world infrastructure network of Fig. 4.3. The methodology is based on the classical theory of connectivity and reliability polynomials. Application servers in a network provide services for users via network connections. One application consists of one or more services and applications may be duplicated or multiplied in the network. Networked services can be included in the model by aggregating networked services with the physical network layer.

Network resilience is modelled with classical reliability polynomials (Ball et al. 1995) and the topological structure of the network. We provide examples of mathematical methods for calculating network resilience, availability of services, and overall utility of networked services. These methods can be used for modelling technical failures, cyber-attacks against systems and physical attacks against the infrastructure. At the physical network level, we define resilience $p_{s,t}$ as the probability of functioning connection between nodes s and t (For example, node s is a user and node t is a service).

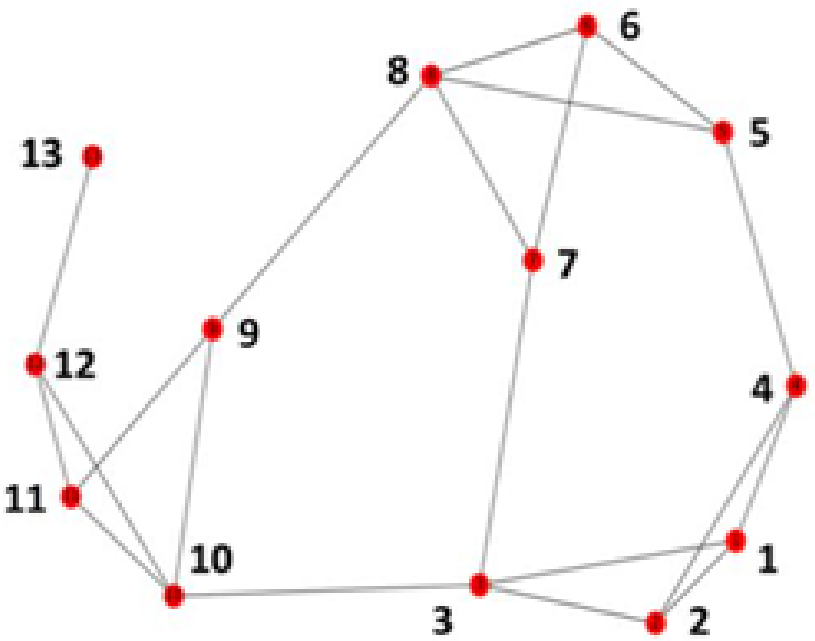

Fig. 4.3. Infrastructure network topology of 13 nodes and 20 bidirectional links (40 directed links) used to demonstrate mathematical methods in the text. Communicating devices are modelled as nodes, and connections between the devices are modelled as links between the nodes.

We use the methodology of Section VI in (Kuikka 2019b) and (Kuikka & Syrjänen 2019) for combining physical network and service layers: connectivity on the physical network level and usability from a user's point of view in performing military tasks or operations. These two aspects can be modelled separately and presented as two matrices. Mathematically, these two matrices are combined with the Hadamard matrix product. The Hadamard product is the entry-wise product of two matrices of the same dimension. The service layers describe the value of functionalities from a user's point of view between nodes that may be located far away in the physical network structure.

Services are evaluated individually for all of the connections between all pairs of nodes in the networked services regardless of the number of physical links in the communication network between any two given nodes. This method enables describing various uses of different services and operational situations. Practical examples with numerical results are presented in (Kuikka & Syrjänen 2019). Different methods for computing probability of connectivity are available but adjacency matrices (Newman 2010) are illustrative and computationally efficient for small and medium-sized network structures in the military context.

After aggregating the physical network level and the service level, the utility of networked services can be analysed. In a military context, this information can be used for planning more resilient and robust systems. Also, critical components can be identified, for example, by investigating the effects of removing a node from the modelled system. In the following, we provide a brief description of the mathematical method

used in aggregating the communication level and the service level of networked systems (see Fig. 4.4).

First, the reliability matrix between every pair of nodes in the network is computed. In this context, the reliability matrix is defined as matrix $\mathbb{P}$, whose elements $p_{s,t}$ $, s, t = 1, .., N$ are probabilities of operational connection from node s to t, where N is the number of nodes in the network. To calculate the reliability matrix $\mathbb{P}$, simulations from the reliability values of links (or nodes) are computed. Link reliability values (link weights) are probabilities of operational links used as input values for computing the reliability matrix. Simple examples have been given in Eqs. (69) – (72), where we have used the notation $p_{s-t} \equiv p_{s,t}$. A pseudo algorithm for computing the elements of network reliability matrix $\mathbb{P}$ is presented in (Kuikka & Syrjänen 2019).

After evaluating the matrix $\mathbb{P}$, the respective use of connections and services are evaluated. These matrix elements are denoted by $C_{s,t}$ $, s, t = 1, .., N$. The values of $C_{s,t}$ are obtained by aggregating the link weights between the common nodes across of each service and normalising the values by using the number of services N_S as the divider (Kuikka & Syrjänen 2019):

$$\mathbb{C} = \frac{1}{N_S} \sum_{z=1}^{N_S} \mathbb{S}_z. \tag{64}$$

Matrix $\mathbb{S}_z$ describes the usability of service between all pairs of nodes in the network. The simplified aggregation method of Equation (64) is used because the services and their weighting are always considered in the same context of military operations as opposed to cases where entirely different systems are evaluated together.

The values of $\mathbb{C}$ are used as input for computing the overall utility $U_{s,t}$ $, s, t = 1, .., N$. We define matrix $\mathbb{U}$ by taking the Hadamard product of matrices $\mathbb{C}$ and $\mathbb{P}$ (Kuikka & Syrjänen 2019):

$$\mathbb{U} = \mathbb{C} \circ \mathbb{P}. \tag{65}$$

The definitions of $\mathbb{C}$ and $\mathbb{U}$ adapt to various practical situations and uses of services. This is a consequence of the fact that the physical network infrastructure and functional usefulness of services are separated and can be evaluated or modelled independently. Numerical examples are provided in (Kuikka & Syrjänen 2019).

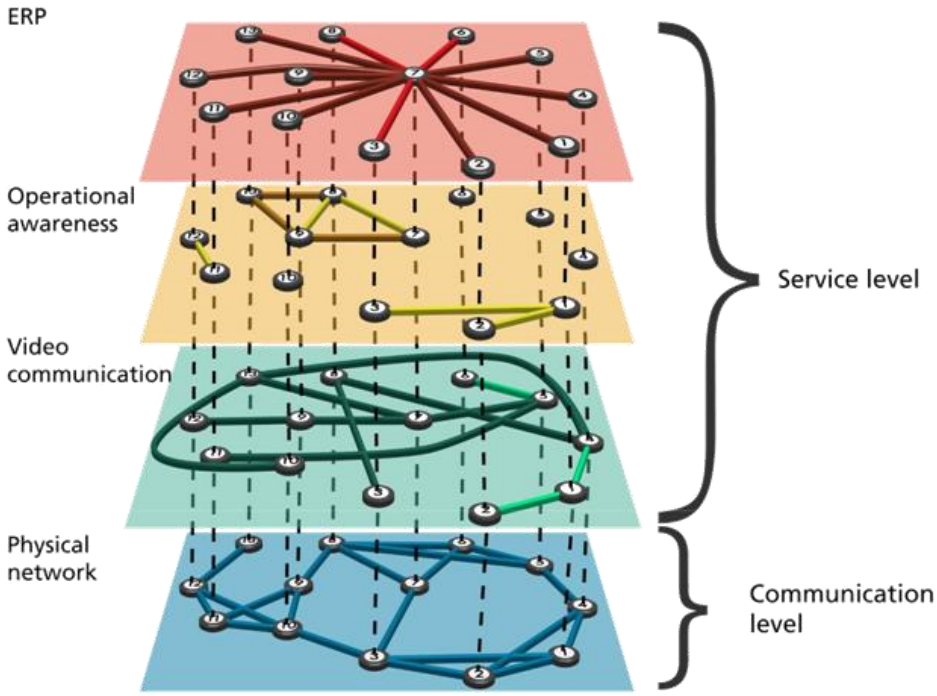

Fig. 4.4. Communication level representing the physical network and the service level illustrating the different services required between the nodes.

4.3. Optimal Sensor Placement

Methods of complex networks analysis (Newman 2010; Newman 2012; Barabási 2016) and network propagation modelling (Kuikka 2018a; Kuikka 2021) are used to study communication networks. We present a method for resolving the optimal placement of sensors in general network topology (P VII). In this context, a 'sensor' is defined as a device or logic to monitor and control Internet traffic. Sensor placement is calculated by balancing maximised closeness to the monitored nodes with sensor location distribution for protection. The model serves cyber defence planning and critical infrastructure defence. The model uses node and link weights as parameters to describe information quantity, data importance to a defender, and criticality to the network operator. General network analysis, together with a penalty factor for describing distributed defence requirements, underpins the methodology.

The aspects of centrality (Gómez 2019) and network resilience (P VII) are considered equally. Penalty components are weighted with a tuneable parameter, and they are introduced to increase the defensive capability against targeted attacks. With low parameter values, sensors are placed in more central nodes close to each other. With high parameter values, sensors are placed on resilient locations. Adjusting both the link weights and the penalty parameter illustrates that the distribution of sensors can be either at the crucial crossroads in the network or widely separated from each other in the network structure. The parameter adjustment determines the level of protection. The applicability of the method is illustrated with three real-world networks, and the results are both intuitive and practical.

Various types of sensors exist in modern communication network infrastructures and architectures. Sensors are utilised for both monitoring and controlling. Different technical solutions have been designed for cyber defence and traditional management functions. In cyber defence, sensor systems can be utilised to mitigate malicious actions and software or even hostile actors themselves. The method can be used for determining optimal sensor placement in a communication network at both the physical and logical levels. Physical topology is a map of network structure, and logical topology illustrates how services are used within the network. The model has node- and link-weights as parameters to describe the value of data and services in the nodes and information transfer via the links.

In the methodology of complex network analysis (Newman 2010), the method for optimal sensor placement can be based on two alternative basic measures: closeness centrality or betweenness centrality (Kuikka 2018a; Kuikka 2021). Closeness centrality indicates how close a node is to all other nodes in the network. Nodes with high betweenness have considerable influence within a network by their control over information passing between other nodes. The two measures for optimal sensor placements are the corresponding sums of the influence measures of nodes with an extra parameter where the influence between the sensors is eliminated. The extra component is multiplied by a tunable parameter that acts as a penalty element in the model. The penalty parameter is incorporated in the model to increase the capability to defend against targeted attacks. It is crucial that the model is 'global' and takes into account all the paths in the network to make it possible to calculate the penalty component. The model also considers information flow direction and both link and node weights. The weights can be applied to model information quantity and data importance to a defender.

Next, we write an iterative formula for calculating the spreading probabilities in networks, by considering paths between a source node and a target node (Kuikka 2018a; Kuikka 2021). In the context of communication networks, the model describes information flow. In the model, links are directed and both links and nodes can have weighing factors. It is left up to user discretion to plan and select weights.

In step i paths from source node s to target node t are combined iteratively in the descending order of $L(i,3)$, the common path length before branching off the paths as (Kuikka 2018a)

$$P_{i,L(i,2)} = P_{i,L(i,1)} + P_{i-1,L(i-1,2)} - \frac{P_{i,L(i,1)}P_{i-1,L(i-1,2)}}{W_{L(i,3)}}, i = 1, \dots N_L - 1. \quad (66)$$

The path length $L(i,2)$ in iteration i is $L(i,2) = max\{L(i,1), L(i-1,2)\}$ and the common path length of $L(i,1)$ and $L(i-1,2)$ is denoted by $L(i,3)$. The number of different paths from the source node to the target node is denoted by N_L. The iteration starts with the two paths with $P_{1,L(1,1)} = W_{L(1,1)}$ and $P_{0,L(0,2)} = W_{L(0,2)}$

having the longest common path length $L(1,3)$. If there are more than two paths with the same common path length, these paths can be processed in any order. In the subsequent steps of the iteration, combined paths are processed in the same way as the original paths of the network. A numerical example in (Kuikka 2018a) illustrates the algorithm in practice. The probability of information propagation between the two nodes is the final result of the algorithm after all the paths have been processed. Thus, for source node s and target node t we have $C(s,t) = P_{N_L-1,L(N_L-1,2)}$. In the last step of the iteration the length of the last combined path is $L(N_L - 1,2)$.

We use the following objective function (P VII) for the measure of sensor placement:

$$M_c(\pi_c) = \sum_{\substack{s\notin\mathcal{S} \text{ and } t\in\mathcal{S} \text{ or} \\ s\in\mathcal{S} \text{ and } t\notin\mathcal{S}}} C(s,t) - \pi_c \sum_{s\in\mathcal{S} \text{ and } t\in\mathcal{S}} C(s,t). \tag{67}$$

The first summation (component) is taken over sensor/non-sensor pairs of indexes and the second summation over sensor pairs of indexes. The set of nodes equipped with sensors are denoted by $\mathcal{S}$. In the first summation, index pairs between non-sensor nodes and between sensor nodes are not included. The second part is the penalty component multiplied by the penalty parameter π_c. The measure $M_c(\pi_c)$ is maximised in the procedure for determining the optimal locations of the desired number of sensors. Later in this article, all the numerical results are calculated by $M_c(\pi_c)$.

The first component in $M_c(\pi_c)$ is based on the closeness centrality measure defined in (Kuikka 2018a; Kuikka 2021). As mentioned earlier, an alternative measure M_b in (P VII) is based on the betweenness measure of (Kuikka 2018a; Kuikka 2021)

$$M_b(\pi_b) = \sum_{n\in\mathcal{S}} b_n - \pi_b \sum_{s\in\mathcal{S} \text{ and } t\in\mathcal{S}} C(s,t)\,, \tag{68}$$

where

$$b_n = \frac{\mathcal{C} - B_n}{\mathcal{C}} \text{ and } \mathcal{C} = \sum_{s,t\in All\ nodes} C(s,t). \tag{69}$$

The quantity b_n is the betweenness centrality value of node n. The numerical value of B_n is equivalent to the closeness centrality measure when node n is removed from the network. It is to be expected that in most cases the optimal values of measures $M_c(\pi_c)$ and $M_b(\pi_b)$ are close to each other because normally central nodes have also a high betweenness value. With atypical network topologies or system configurations, the two objective functions can provide different results. A semantic difference between the quantities is that $M_c(\pi_c)$ measures access to information and $M_b(\pi_b)$ emphasises information flow between the nodes in the network. These points of view may be different in planning backbone infrastructure networks, local networks, or embedded system structures.

In the literature (Newman 2010; Martin et al. 2014), most of the network models include only local interactions. For this reason, they are incapable of describing long-distance influence between nodes. The network model in this study (P VII) includes all interactions between the nodes of a network, that is, short and long-distance connections, enabling the calculation of the penalty component in $M_c(\pi_c)$ and $M_b(\pi_b)$.

Three different real-world network topologies are used for demonstrating the model (P VII). As an introduction, a model with a simple physical network structure of Fig. 4.3 is presented. Additionally, the use of a service layer on top of the physical network with the network of Fig. 4.4 is illustrated. The second example network in (P VII) describes the same network as the first network, but the complex structure is shown in more detail. To prove the usability of the methodology in a more realistic situation, where the number of sensors is higher, calculations are performed with the Sprint operator network in the USA is analysed in (P VII).

The essential question is which model parameter values should be used in the calculations. The number of sensors is determined by the managers or operators of the network. The decision depends on the cost of the technology and the number of available resources for maintaining the sensor system. The optimal placement of sensors has a clear pattern with a wide range of model parameters. This is a favourable feature of the model because results are not highly sensitive to the model parameters. For example, it may be difficult to estimate the best numerical value for the link weights. Coarse estimations and experimenting with a few link weight values are sufficient.

First, we discuss the results of Table 4.2 on the left side for the communication layer of Fig. 4.3. When only one sensor is placed in the network, node 3 is optimal. Further, nodes 3 and 8 are optimal for two sensors. Different configurations for three sensors are suggested by the analysis: nodes 1, 8 and 10 are optimal for low activity events with low requirements for protection and 1, 8 and 12 for high protection. The corresponding configurations are 3, 8 and 10 (1, 6 and 13) for moderate activity events with low (high) protection. The final selection can be made among the suggested nodes, calculated with different model parameters, and judging with possible additional budgetary and system management requirements.

The right side of Table 4.2 shows an example of sensor placement in the network topology of Fig. 4.3, where the service layers of Fig. 4.4 are also considered. The results are different because services carry a high weight in the calculations. If the physical network layer, or the availability of services, were more important, there would be more results similar to those on the left side of the table.

Services are essential from the users' point of view. Consequently, the sensor placement should consider not just the physical network structure but also the networked

services. Services can be categorised and rated based on the value of cyber and operational functionalities. In the military context, a capability is based on the accessibility and usability of these functionalities. In real-world applications, the value of services depends on the task and operation of the scenario.

Table 4.2. Optimal sensor placements computed for the network topology of Fig. 4.3. Links weights, penalty parameter values, the number of sensors, and optimal nodes are denoted by W_l, π_c , #, and locations of sensors correspondingly. The results are for the communication layer and both communication and service layers of Fig. 4.4.

Communication Layer in Fig. 4.3										**Communication and Service Layers in Fig. 4.4**			
W_l	π_c	#	**Sensors**		W_l	π_c	#	**Sensors**		W_l	π_c	#	**Sensor**
0.1	1	1	3		0.5	1	1	3		0.1	1	1	7
		2	3,8				2	3,8				2	7,11
		3	1,8,10				3	3,8,10				3	1,9,13
		4	3,4,8,11				4	1,6,8,10				4	1,5,9,13
		5	3,4,6,9,12				5	1,3,6,8,11				5	1,5,8,9,13
	5	1	3			5	1	3		0.1	5	1	7
		2	3,8				2	4,11				2	1,9
		3	1,8,12				3	1,6,13				3	1,9,13
		4	3,4,8,12				4	1,6,11,13				4	1,5,9,13

5.

RESULTS

Our aim has been to develop a common modelling framework and methodology in the areas of military capabilities, combat outcomes and operability of networked systems. Our solution is to use probabilistic modelling as a common methodology in developing quantitative models in these areas. As the three topics are all related to each other, we have created a collection of articles that define models in these areas (P I – P VII). We have proposed novel mathematical models and corresponding formulas in all these areas (see Fig. 1.1). In this section, we summarise the main results.

5.1. Summary of the Results

In this section, we sum up the content of this study. In Section 6, we will provide a wrap-up of how the research questions have been answered. The primary research question in Section 1.5 asks how to describe central military concepts through quantitative models. Our solution to the research question is to apply the methods of probability theory to all of the select modelling areas. This is the way of developing models that use both quantifiable predictive variables and have quantifiable model outcomes. Very detailed and complicated models can be constructed by probabilistic methods but we keep our approach basic, still developing useful models that can realistically describe the concepts in the desired level of abstraction or scope.

The main research question has been discussed throughout Sections 2, 3, 4 and also in this section. Some examples of the interrelationships of modelling military capabilities, combat modelling and modelling networked systems have been summarised in Table 1.1. Representative practical examples of all the mathematical models have been presented in Sections 2, 3 and 4 and more examples can be found in the original papers (P I – P VII).

Research Questions 1.1-1.3 (see Section 1.5)
In capability modelling, a conceptual specification of capability areas and their hierarchical sub-level structure is assumed to be available. The model describing military capabilities (P I, P II, P III) is based on a system of systems approach where functional high-level capabilities are decomposed into system capabilities and sub-capabilities. We present a theoretical procedure for modelling system capabilities or capabilities of sub-system functionalities. The procedure is based on decomposing high-level capabilities into low-level serial and parallel sub-capabilities classified as necessary and alternative functionalities. One system can have functionalities (see Fig. 2.3) in several capability areas. A practical example of three capability areas, two systems and three scenarios has been presented in the articles (P I, P II, P III). Building a very granular

structure for functionalities is complex because of the increasing number of variables and the difficulty of defining various functionalities as a structure of serial and parallel components. However, incorporating a limited number of additional systems can be easier due to the iterative procedure of adding a system to the existing set of systems (see Section 2.3).

As we assume that adding or removing systems from the ensemble of a system of systems does not affect the remaining system capabilities, recalculation of the entire model is not needed. This is dissimilar to many other models like the Analytical Hierarchy Process (Saaty 1990). In fact, the idea of our model is to isolate system capabilities from the effects of other systems. Dependence on the scenario remains because of the probabilistic definition of capability. In the procedure, expert evaluations of the effects of deploying a new system are conducted at the high functional level of capability areas for the entire system of systems. Also, we present mathematical methods to calculate the number of system units by optimising capability requirements through multiple system capabilities (P III).

Research Questions 2.1 and 2.2 (see Section 1.5)
Performance in combat activities or warfare operations and military capabilities are two sides of a coin. Both of the modelling approaches for military capabilities and combat performance are based on probabilistic modelling. A novel analytical form of a combat equation has been presented in (P IV, P V). The equation has been derived assuming that attrition processes obey geometric Brownian motion of stochastic analysis. Two variants for modelling stopping rules have been presented, both of them giving similar results for the probability of victory in a battle between the attacker and the defender. In the second variant, the stopping rules are modelled as constants as a function of time. The corresponding attrition process is defined as the difference of individual attrition processes of the two opposing sides.

The second version of the model in (P V) makes it possible to derive a formula for the expected duration of a specific battle in an analytical form. Combining the results for victory and expected duration of battles enables better validation of the model, more accurate evaluation of model parameter values, and discovering new properties of combats. First, we conclude that the model predictions of probabilities and expected durations are in line with the empirical observations. After that, assuming that the model describes combats sufficiently well, we have discovered that in asymmetrical attrition processes (having high values of the ratio S of Equation (40)), asymmetrical decision boundaries are needed to explain the modelling results. Not higher variances alone, nor wider decision boundaries alone, can explain the results of the model. This is a reasonable outcome describing a resilient opposing force in an asymmetrical attrition process.

Research Questions 3.1-3.2 (see Section 1.5)
The operability of networked systems is related to military capabilities as specifications of military capabilities typically determine network operations as one of the main capability areas (Kuikka 2020). In modern warfare, communication and network systems are increasingly important. The use of sophisticated technological systems is an advantage in combat especially when one of the opposing sides is superior in the use of technology. Different topics concerning networks are discussed in this study or related conference articles by the same author: resilience of networked systems (P VI, P VII), optimal sensor placement in network structure (P VII) and leadership in military organisations (Kuikka & Nikkarila 2019). Technical networks and social networks are investigated with two different network models designed for describing connectivity (Ball et al. 1995) in physical networks and information or influence propagation (Kuikka 2018a; Kuikka 2021) in social networks respectively.

(P VII) introduces a method for calculating optimal sensor placement in the network topology. Sensors are placed on optimal and resilient locations to monitor network traffic or control activities between network segments, network devices, information services or systems' users. Optimal placement of sensors is important when a limited number of sensors are positioned in the communication network infrastructure. The novelty of the methodology is that network structures, the value of information services, the importance of network connections, and protection aspects are considered consistently through distributed sensor placements in the same model.

5.2. Validation of the Models

According to (Jaiswal 1997; Gass 1983) face validity or expert opinion, sensitivity analysis and hypothesis validity apply to model validation in military systems (see Section 1.6).

The face validity or the expert opinion has been inspected in peer-review processes of the journals where the articles of this study have been published. All the models have been published in journals on the fields of military operations research or operations research.

Variable and parameter validity or sensibility analysis has been accomplished by demonstrating all the models with numerical empirical data, or at least, with realistic numerical parameter values. These use cases have also been validated in the journal approval processes.

Hypothesis validity of the capability models (P I, P II, P III) is straightforward because the model structure reflects interrelations between sub-capabilities or system capabilities with capability areas. UAVs and satellites are used in the example case study. Results in (P IV, P V) have been compared with historical combat data (Hartley

2001) and with other results from analytical combat equations (Hartley 2001; Helmbold 1987; Willard 1962). Two versions of the model have been developed in (P VI) to validate the modelling approach of using Markov processes in describing state transitions of networked military services. Results in (P VII) have been compared with the model in (Alenazi 2018).

5.3. Reliability in Model Predictions

According to (Abrahamsson 2002) five kinds of uncertainties of reliability in model predictions can exist: specification of the problem, conceptual modelling, computational implementation, input parameters and calculations with documentation of the results (see Section 1.6). Next, we discuss these aspects in the context of this study, which mainly concerns capability modelling. Similar sources of uncertainties exist in the modelling of combat effectiveness or outcomes and networked military systems.

In the capability modelling, we have delimited conceptual analysis outside this work because several conceptual models of military capabilities are available, for example in (Kosola 2013; JETCD 2009). On the other hand, this is a major issue concerning reliability as we have pointed out in (Kuikka & Peltotalo 2020). Our articles in (P I, P II, P III) are based on two questionnaires where a common understanding of the relevant concepts was required. Different perceptions and differing opinions about the meaning of concepts can also undermine the interpretation of response variable values in the modelling results.

Respondents had an opportunity to discuss and debate upon the subject and related concepts (this has some resemblance with the Delphi method, see Section 6). The limited time to answer the questions reduced the authority, personality, or reputation of some participants from dominating others in the process. An additional source of uncertainty is the bias based on human judgement. Systematic and cognitive biases have also been detected: Events that are vivid, recent, unusual or highlighted by the media are assigned high probabilities (Goodwin & Wright 2010).

One counterbalancing point of view in the application of technology forecasting is that the time horizon is 10 or 20 years. Uncertainties in technological development are considerable and even the concepts themselves can change in the future. The concept of protection capability is particularly multifaceted (Kuikka 2019a). Higher standard deviations of the numerical values in the questionnaire and somewhat abnormal modelling results may reflect this uncertainty.

In the case of protection capability, we discussed in Section 2.4 the possibility of incomplete modelling when the cooperative use of two or more systems provide more (or less) capability because of reinforcement effects. This is an example of a modelling error where the model does not describe real-world events accurately because the

model is not detailed enough. A more detailed model describing protection capability in the case of UAVs and satellites is presented in Section 2.4. In this particular case, we encounter a situation that demonstrates that sources of uncertainty are not independent. In the original questionnaire, no questions were asked whether an extra capability can be gained through simultaneous deployment of systems due to their mutual intensifying effects.

We conclude that the arrangement of data collection, granularity of the model, conceptual modelling etc. should be designed with the same principles in mind. In our case, capability modelling and combat modelling have been conducted on a highly conceptual level. Also, the application area of capability modelling has been long term technology forecasting. In combat modelling (P IV), the methods are limited by the quality and information content of the available empirical combat data (Hartley 2001). Detailed mathematical models for describing networked military systems are feasible but obtaining the values of input parameter values may be a problem. Including networked services involves evaluation and quantification of the importance or utility of services in military operations. This kind of evaluations may be difficult to accomplish and they are often based on human judgment in a specific scenario.

6.

SUMMARY – ANSWERING THE RESEARCH QUESTIONS

In summary, we have presented modelling methods in three intertwined areas of defence studies: military capabilities, combat outcomes and utility of networked services. Several quantitative measures for evaluating the effectiveness and capabilities of different system configurations are proposed that can be used in planning defence systems. The models are useful for both explanatory and predictive purposes of the concepts in the military domain. We provide practical use cases for all the methods and models proposed in this study.

The solution to the main research question of using probabilistic methods has been endorsed throughout this study by presenting probabilistic models in different areas of military studies. The solution to the research question has been demonstrated with practical examples by showing that the methods are useful and they provide reasonable and consistent results (P I – P VII). Research sub-questions (see Section 1.5) are listed in Table 5.1 with a column indicating corresponding original articles where the sub-questions have been addressed. The most important equations formulated in this study answering the sub-questions are indicated in the last column. Highlights of the primary research question and the seven sub-questions are listed after the table with short answers corresponding to the questions.

In Section 1.4, we discussed research gaps in the existing literature in the areas of generic quantitative modelling of military concepts, modelling of military capabilities, modelling of combat outcomes and modelling of networked systems. Corresponding research questions on research objectives were discussed in Section 1.5.

Table 5.1. List of research gaps (RG), research sub-questions (RQ) and articles (P) where they have been addressed. The most important model equations of this study are indicated in the last column.

RG	RQ	P	Short description of the context in articles	Eqs.
**)	*)	all	Common methodology for describing central military concepts	all
1	1.1	I	Capability areas and system capabilities	8,9
	1.2	II	Technology areas and technological development	2
	1.3	III	Optimised number of system units, mathematical methods	24-28
2	2.1	IV	Combat equation, Probability to win a battle	39-45,52
	2.2	V	The mathematical formula for the expected duration of a battle	53
3	3.1	VI	Networked systems modelled with the Markov matrix formalism	60-63
	3.2	VII	Optimal sensor placement in the network structure	66-69

*) **The main research question:** The answer to the main research question is answered through answering the sub-questions 1.1 – 3.2. Probabilistic modelling provides a methodology for describing all the sub-questions with common quantitative methods and concepts.

**) The main research gap corresponding to the main research question was discussed in Section 1.4.

Research question 1.1:
Research question 1.1 has been answered in Equations (8) and (9) and explained in (P I). An iterative procedure (Kuikka 2019a) is presented for multiple sub-capabilities composed of serial and parallel structures. Capabilities and sub-capabilities are produced by systems and system functionalities. Equations (3) – (7) show an example for two systems.

Research question 1.2:
Research question 1.2 is answered in Equation (2) and explained in (P II). The method takes into account the interdependencies of different development levels between technology areas.

Research question 1.3: Research question 1.3 is answered in Equations (24) – (28) and explained in (P III). This is an extension to research question 1.1 where the number of system units was not considered. The general method is demonstrated with a numerical example of two systems in (P III).

Research question 2.1: Research question 2.1 is answered in Equations (39) – (45) and Equation (52) and explained in (P IV). Equations (39) and (43) show probabilities for the attacker to win a battle and the defender to lose a battle, respectively. In the analysis, these two quantities should give similar numerical values. Demonstrations on how to use the model are provided in (P IV). Equation (52) corresponds to a modified definition of stopping rules (boundaries) of the combat. Because these two versions of the model provide similar results for the probabilities of winning combats, the credibility of the modelling method is higher.

Research question 2.2: Research question 2.2 is answered in Equation (53) and explained in (P V). Expected duration is computed with the second version of stopping rules and the corresponding definition of the attrition process for the combat explained in (P V).

Research question 3.1: Research question 3.1 is answered in Equations (60) – (63) and explained in (P VI). Two different models for modelling networked systems' operability demonstrate how Markov matrix formalism can be applied for modelling the resilience of networks and services. More examples of how to model networked systems, including both technical and human resources, are provided in the references.

Research question 3.2: Research question 3.2 is answered in Equations (66) – (69) and explained in (P VII). The article introduces an optimal sensor placement method

that considers monitoring and controlling processes, the resiliency of the network, networked services and the structure of the entire network.

We have conducted literature reviews and discovered research gaps where more research is needed. On this basis, we have formulated research questions: one primary research question and seven sub-questions. Answers to the research questions have been provided in the original journal articles (P I – P VII). In this work, we have introduced the research articles and discussed common characteristics and interrelations between them.

7.

FUTURE RESEARCH

The proposed methods and the general methodology of this study create various opportunities for future research. One of the goals has been to build a foundation for developing new quantitative models in the military domain. Next, we discuss some possibilities in detail. Probably several additional cases exist where the proposed methods can be applied. This is possible because the methods are based on the common methodology of probabilistic modelling.

Future Research of Military Capabilities (see Section 2)
The example cases of this study consider only a limited number of capabilities and systems. It is therefore evident that the next phase of implementing the method is to include more capabilities, systems and their sub-structure. An interesting research direction would be to consider hierarchical structures of capabilities with the structures of a system of systems in cases where the two structures are not similar, that is, the system structure is not hierarchical. As explained in Sections 2, the structure of a conceptual capability model (Kosola 2013) is hierarchical, and the structure of systems and sub-systems is a network of serial and parallel components (P I, P II, P III). Also, more research is needed in cases where the assumption of serial and parallel functionalities does not apply perfectly. One example of this was discussed in Section 2.4.

In this study, the main application of modelling military capabilities was in technology forecasting and technological development (see Sections 1.2.1 and 2). One idea for future work is to promote our method as a new tool for technology forecasting, also in non-military areas. Our method has a resemblance to the well-known Delphi method. The Delphi method (Helmer 1967) is a process used to arrive at a group opinion or decision by surveying a panel of experts. The main difference between the two methods is that our method considers interdependencies of capabilities or systems and simultaneous use of systems in operations.

Another method commonly used for organising and analysing complex decisions is the Analytic Hierarchy Process (AHP). AHP (Saaty 1990) is based on making a series of judgments of pairwise comparisons between alternatives. As explained in Section 2, systems are not compared pairwise which makes our method easier to use because the number of evaluations is smaller. This is emphasised when the number of alternatives is higher. AHP is one of the methods in the area of multiple-criteria decision analysis (MCDA), that explicitly evaluates multiple conflicting criteria in decision making (DCLG 2009; Tzeng & Shen 2017). There is also a wide range of research in the fields of quantifying uncertainty and evaluating probabilities of human judgment and systems engineering (Parnell 2017). Structural equation modelling (SEM) can impute

relationships between unobserved constructs and observable variables (Cheung 2015). Comparing our method with Delphi, AHP, SEM or other MCDA methods would be an interesting topic for a research article.

Research question 1.3 in Section 1.5 has been answered in (P 3) and demonstrated with the example use case of UAVs and satellites. The method solves the problem of calculating the optimal number of system units within a predefined budgetary constraint (see Section 2.5). This application adds value to the methods of modelling capabilities in Sections 2.1 – 2.4. To the knowledge of the author of this study, the combined use of capability modelling for technology forecasting and calculating the optimal number of system units is unique in the literature. The actual implementation of the methods in a real-world implementation would deliver on the promises of operations research.

Future Research of Combat Effectiveness (see Section 3)
The method of modelling combat outcomes and combat durations is demonstrated with the historical combat data from (Hartley 2001). Combat effectiveness as a function of time has been studied in (P IV) in the three cases where empirical time-dependent data is available. This research could be extended to investigate how the empirical volatility value can be used as a model parameter. Also, the present model could be modified to consider the combat as a chain of successive phases of one combat and not as separate daily battles (P IV).

The question arises whether the model applies to modern conflicts which are characterised by complex scenarios, weaponry, tactics and various technical systems (Hartley 2020). In (P IV, P V) combat manpower, armoured personnel carriers (APC), artillery and tanks were included in the model by weighting them by 1, 5, 20 and 40, respectively. An alternative idea would be to refine the model for stopping rules to take into account the technical systems of the two opposing sides. Because the method in (P IV, P V) is designed for modelling macroscopic level phenomena of combats, developing detailed models on this basis may be limited. Systems engineering (Parnell 2017) methods can be useful, as such, or as sub-models of our model. Along these lines, using the traditional Quantified Judgment Model (QJM) could be exploited as a sub-model for the models in (P IV, P V). QJM takes into account, in addition to weaponry and force sizes, human factors such as leadership, troop morale and operative factors like terrain and weather (Dupuy 1985).

Future Research of Networked Military Systems (see Section 4)
The first example use case involves applying standard Markov processes (Florescu 2015) to describe transitions between failed and operative states of networked systems. The Markov matrix formalism can be used to model different system arrangements composed of network links and nodes in different states. Modern computing systems facilitate calculations of large system configurations.

The second example use case applies the information spreading model (Kuikka 2018a; Kuikka 2021) and a procedure for placing sensors optimally in a network structure. The optimisation procedure considers two opposing requirements: the centrality of sensors and protection against failed connections in the network structure. From the modelling perspective, using different network models instead of (Kuikka 2018a; Kuikka 2021) would be an interesting exercise. The standard network reliability theory (Colbourn 1987; Ball et al. 1995) is a choice for modelling network connectivity. As we expect similar results, the network connectivity model would serve mostly for the sensibility analysis of the sensor placement method.

The functional form of the objective function can have different representations in specific circumstances. One alternative for the objective functions is provided in (P VII) emphasising the betweenness property (Equations (68) and (69) in Section 4.3) instead of the centrality (Equation (67) in Section 4.3) in optimising the reachability of controlled or monitored nodes from the sensor locations. Also, detailed modelling can be attained by considering individual attributes of nodes and links in the network. For example, network nodes providing important services can be weighted in the network structure. Both network models of (Kuikka 2018a; Kuikka 2021) and (Colbourn 1987; Ball et al. 1995) can use node and link weights as input parameters. The importance of different services can be evaluated based on their military capability value which they provide to the end-users.

In general, considering the services produced by information and application servers located in the network structure is a topic for future research. This kind of research may be more important in military applications than in non-military environments. The reason for this is that adversarial attacks are possible in military environments. Protecting services and connections and multiplying critical servers and connections are typical actions in military applications. Mathematical models are needed in optimising these protective measures.

8.

CONCLUSIONS

The main research question was asked at the beginning of this study: "How to describe central military concepts through novel quantitative models to support planning and allocation of networked system capabilities." The requirement of quantitative models naturally leads to probabilistic methods in mathematics. Using common concepts and methods enables comparing the modelling results. This approach solves the old problem of comparing different military capabilities with each other. Using probabilistic modelling answers the generic research question and the first sub-question of how to model military capabilities with quantitative methods. Utilising probabilistic methods solves also the second and third sub-questions. In summary, the answer to the main research question is by developing probabilistic models of military concepts.

We have proposed novel probabilistic methods for modelling military capabilities and combat outcomes. The proposed methodology fills the research gaps we have discovered in the existing literature of military operations research. Lastly, we have shown how the operability of networked systems can be subsumed in the modelling framework. Research gaps, research questions and answers to the questions have been summarised shortly in Section 7.

In this study, we have covered a wide range of military operations with the three use cases of military capabilities, combat outcomes and networked systems. We have discussed the results of this study and commented on related issues of validity and reliability in model predictions. Possible future research has been discussed. The proposed methods of probabilistic modelling can be applied to many other problems and capability areas in the military domain. The methodology of using probabilistic modelling provides decision-makers with a quantitative basis for planning and allocation of military capabilities.

REFERENCES

Abrahamsson, M., 2002. Uncertainty in Quantitative Risk Analysis – Characterisation and Methods of Treatment. Licentiate thesis, Lund University Centre for Risk Assessment and Management.

Alenazi, M.J.F., 2018. On SDN Controller Placement to Achieve Robustness Against Targeted Attacks. In: Cherifi C., Cherifi H., Karsai M., Musolesi M., eds. Complex Networks and their Applications VI. Complex Networks 2017. Studies in Computational Intelligence, vol. 689. Springer, Cham. https://doi.org/10.1007/978-3-319-72150-7_51

Ancker, C.J., Jr., Gafarian, A.V., 1992. Modern Combat Models, a Critique of their Foundations, Published by Operations Research Society of America. Linthicum, MD.

Andersson, K., 2018. On the Military Utility of Spectral Design in Signature Management: A Systems Approach. National Defence University, Finland Series 1: Research Publications No. 21, Doctoral dissertation.

Andersson, K., Bang, M., Marcus, C., Persson, B., Sturesson, P., Jensen, E., Hult, G., 2015. Military Utility: A Proposed Concept to Support Decision-making. Technology in Society, 43, pp. 23-32. https://doi.org/10.1016/j.techsoc.2015.07.001

Armstrong, M.J., 2005. A Stochastic Salvo Model for Naval Surface Combat, Operations Research, 53(5), pp. 830-841.

ATP, 2016. Infantry Platoon and Squad. Army Techniques Publication ATP 3-21.8.

Ball, M.O., Colbourn C.J., Provan, J.S., 1995. Network Reliability. In: Ball, M.O., Magnanti, T.L., Monma, C.L., Nemhauser, G.L., eds. Handbooks in Operations Research and Management Science, Network Models, vol. 7, Elsevier B.V. Netherlands. pp. 673-762.

Barabási, A.-L., 2016. Network Science. Cambridge University Press, United Kingdom.

Bigdeli, A., 2012. Robustness and Vulnerability Design for Autonomic Management. Doctoral thesis, Graduate Department of Electrical and Computer Engineering Department, University of Toronto.

Bigdeli, A., Tizghadam, A., Leon-Garcia, A., 2009. Comparison of Network Criticality, Algebraic Connectivity and Other Graph Metrics. Proceedings of the 1st Annual Workshop on Simplifying Complex Network for Practitioners (Simplex'09), Venice Italy, pp. 1-6. https://doi.org/10.1145/1610304.1610308

Biltgen, P.T., 2007. A Methodology for Capability-Based Technology Evaluation for Systems-of-Systems, School of Aerospace Engineering Georgia Institute of Technology May, PhD thesis.

Biltgen, P.T., Mavris, D.N., 2006. A Methodology for Technology Evaluation and Capability Trade-off for Complex System Architectures. 25th International Congress of the Aeronautical Sciences (ICAS 2006), Hamburg, Germany. Accessed 16.4.2021 http://hdl.handle.net/1853/14520

Bonder, S., Farrell, R., eds. 1970. Development of Models for Defense Systems Planning. Report No SRL 2147 TR 70-2, Systems Research Laboratory, The University of Michigan.

Bongers, A., Torres, J.L., 2019. A Bottleneck Combat Model: An Application to the Battle of Thermopylae. Operational Research: An International Journal (2019). https://doi.org/10.1007/s12351-019-00513-0

Booß-Bavnbek, B., Høyrup, J., eds. 2003. Mathematics and War. Birkhäuser, Basel. http://doi.org/10.1007/978-3-0348-8093-0

Bordon, S., Arbour, B., MacLeod, M.R., 2014. A Method for Hierarchically Prioritizing Capabilities with an Application to Military Manned and Unmanned Aerial Vehicles, Journal of Applied Operational Research 6(1), pp. 39-47.

Bracken, J., Kress, M., Rosenthal, R.E., eds. 1995. Warfare Modeling. John Wiley & Sons.

Bracken, J., 1995. Lanchester Models of the Ardennes Campaign. In: Bracken J., Kress M., Rosenthal, R.E., eds. Warfare Modeling. John Wiley & Sons, pp. 417-435.

Bridgeman, D., 2013. Optimization Models for Resource Allocation under Perturbation, Electronic theses and dissertations, University of Louisville.

Carter, M.W., Price, C.C, Rabadi, G., 2019. Operations Research: A Practical Approach, CRC Press, Boca Raton, FL.

Chen, P.S., Chu, P., 2001. Applying Lanchester's Linear Law to Model the Ardennes Campaign, Naval Research Logistics 48(8), pp. 653-661.

Cheung, M.W.-L., 2015. Meta-Analysis: A Structural Equation Modeling Approach, John Wiley & Sons, Ltd.

von Clausewitz, C., 1832. On war. Reprinted by Princeton University Press, 1989.

Coates, J., Glenn, J.C., 2009. Normative Forecasting, In: Glenn, J.C., Gordon, T.J., eds. The Millennium Project, Futures Research Methodology – Version 3.0, AC/UNU Millennium Project.

Colbourn, C.J., 1987. The Combinatorics of Network Reliability. Oxford University Press, New York.

Davis, P.K., 2002. Analytic Architecture for Capabilities Based Planning, Mission-System Analysis and Transformation. National Defense Research Institute RAND report, Accessed 16.4.2021 https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1513.pdf

DCLG, 2009. Multi-criteria Analysis: A Manual, Published by Department for Communities and Local Government, London. Accessed 18.4.2021 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/191506/Mult-crisis_analysis_a_manual.pdf

Denda, R., Banchs, A., Effelsberg, W., 2000. The Fairness Challenge in Computer Networks. In: Crowcroft, J., Roberts, J., Smirnov, M., eds. QofIS 2000, LNCS 1922. Springer-Verlag, Berlin, Heidelberg, pp. 208-220.

DoD, 2020. Military Capability (US DoD Definition) Definition of the Term 'Military Capability'. Documentation of the United States Department of Defense. Accessed 16.4.2021 https://www.militaryfactory.com/dictionary/military-terms-defined.asp?term_id=3357

Dupuy, T.N., 1985. Numbers, Prediction, and War: Using History to Evaluate Combat Factors and Predict the Outcome of Battles. Bobbs-Merrill, Indianapolis.

Edmonds, B., Page, C.L., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Sales, C.M., Ormerod, P., Root, H., Squazzoni, F., 2019. Different Modelling Purposes. Journal of Artificial Societies and Social Simulation 22(3)6. https://doi.org/10.18564/jasss.3993

Epstein, J.M., 2008. 'Why Model?' Journal of Artificial Societies and Social Simulation 11(4)12.

Fiedler, M., 1973. Algebraic Connectivity of Graphs. Czechoslovak Mathematical Journal 23(98), pp. 298-305.

Fleisher, C., Bensoussan, B., 2007. Technological Forecasting. In: Business and Competitive Analysis: Effective Applications of New and Classic Methods. Upper Saddle River, NJ, FT Press.

Florescu, I., 2015. Probability and Stochastic Processes. John Wiley & Sons, Hoboken, New Jersey.

Fowler, B.W., 1995. De Physica Belli: An Introduction to Lanchesterian Attrition Mechanics Part I. Report DMSTTIAC SOAR 95-01.

Fricker, R.D., Jr., 1998. Attrition Models of the Ardennes Campaign. Naval Research Logistics 45(1), pp. 1-22.

GAO, 1985. Measures of Military Capability: A Discussion of their Merits, Limitations and Interrelationships. The U.S. General Accounting Office report.

Gass, S.I., 1983. Feature Article – Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis. Operations Research 31(4), pp. 603-631. https://doi.org/10.1287/opre.31.4.603

Goertz, G., Mahoney, J., 2012. Concepts and Measurement: Ontology and Epistemology. Social Science Information (51)2, pp. 205-216. https://doi.org/10.1177/0539018412437108

Golden, B., 2013. A Unified Formalism for Complex Systems Architecture. École Polytechnique, PhD thesis.

Gómez, S., 2019. Centrality in Networks: Finding the Most Important Nodes. In: Moscato, P., de Vries, N.K., eds. Business and Consumer Analytics: New ideas. Part III, Chapter 8, p. 401.

Goodwin, G.F., Blacksmith, N., 2018. The Science of Teams in the Military: Contribution From Over 60 Years of Research. American Psychologist 73(4), pp. 322-333. https://doi.org/10.1037/amp0000259

Goodwin, P., Wright, G., 2010. The Limits of Forecasting Methods in Anticipating Rare Events. Technological Forecasting and Social Change 77(3), pp. 355-368. https://doi.org/10.1016/j.techfore.2009.10.008

de Gooijer, J.G., Hyndman, R.J., 2006. 25 Years of Time Series Forecasting. International Journal of Forecasting, 22(3), pp. 443-473. https://doi.org/10.1016/j.ijforecast.2006.01.001

Harris, F.W., 1913. How Many Parts to Make at Once. Factory, The Magazine of Management 10(2), pp. 135-136. https://doi.org/10.1287/opre.38.6.947

Hartley III, D.S., 2001. Topics in Operations Research: Predicting Combat Effects. Institute for Operations Research and the Management Sciences, INFORMS.

Hartley III, D.S., 1995. A Mathematical Model of Attrition Data. In: Bracken, J., Kress, M., Rosenthal, R.E., eds. Warfare Modeling, John Wiley & Sons, pp. 443-465.

Hartley III, D.S., 2020. An Ontology of Modern Conflict: Including Conventional Combat and Unconventional Conflict. Springer International Publishing. https://doi.org/10.1007/978-3-030-53214-7

Hartley III, D.S., Helmbold, R.L., 1995. Validating Lanchester's Square Law and Other Attrition Models. In: Bracken, J., Kress, M., Rosenthal, R.E., eds. Warfare Modeling, John Wiley & Sons, pp. 467-491.

Hausken, K., 2000. The Microfoundations of the Lanchester War Equations. Military Operations Research 5(3), pp. 79-99.

Hausken, K., Levitin, G., 2012. Review of Systems Defense and Attack Models. International Journal of Performability Engineering 8(4), pp. 355-366.

Hausken, K., Moxnes, J.F., 2002. Stochastic Conditional and Unconditional Warfare. European Journal of Operations Research vol. 140, pp. 61-87.

Hausken, K., Moxnes, J.F., 2005. Approximations and Empirics for Stochastic War Equations. Naval Research Logistics 52(7), pp. 682-700.

Helmbold, R.L., 1987. Do Battles and Wars Have a Common Relationship Between Casualties and Victory? Technical paper, Presented at the 57th MORS Symposium, Fort Leavenworth, Kansas. Accessed 16.4.20121 dtic.mil

Helmer, O., 1967. Analysis of the Future – The Delphi Method, RAND Corporation Raport. Accessed 16.4.2021 https://www.rand.org/content/dam/rand/pubs /papers/2008/P3558.pdf

Hill, R., McIntyre, G.A., Tighe, T.R., Bullock, R.K., 2003. Some Experiments with Agent-Based Combat Models. Military Operations Research 8(3), pp. 17-28.

Hoppe B., Reinelt C., 2010. Social Network Analysis and the Evaluation of Leadership Networks. The Leadership Quarterly (21), pp. 600-619.

Horne, G.E., Schwierz, K.-P., 2008. Data Farming Around the World Overview. In: Mason, S.J., Hill, R.R., Mönch, L., Rose, O., Jefferson, T., Fowler, J.W., eds. Proceedings of the 2008 Winter Simulation Conference. IEEE: Miami, Florida, pp. 1442-1447.

Hughes, W.E., 1995. A Salvo Model of Warships in Missile Combat Used to Evaluate their Staying Power. Naval Research Logistics 42(3), pp. 267-289.

Inman, L., Anderson, T.R., Harmon, R.R., 2006. Predicting U.S. Jet Fighter Aircraft Introductions From 1944 to 1982: A Dog Fight Between Regression and TFDEA. Technological Forecasting and Social Change 73(9), pp. 1178-1187. http://doi.org/10.1016/j.techfore.2006.05.013

Jaiswal, N.K., 1997. Operations Research in Defense. In: Military Operations Research. International Series in Operations Research & Management Science, vol 5. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-6275-7

Jaiswal, N.K., Nagabhushana, B.S., 1995. Termination Decision Rules in Combat Attrition Models. In: Bracken, J., Kress, M., Rosenthal, R.E., eds. Warfare Modeling, John Wiley & Sons, pp. 273-287.

James, A.D., 2013. Emerging Technologies and Military Capability. RSIS/Policy Briefs, Nanyang Technological University. Accessed 18.4.2021 https://www.rsis.edu.sg/wp-content/uploads/2014/07/PB131101_Emerging_Technologies_and_Military_Capability.pdf

JETCD, 2009. Joint Capability Areas (JCA). Joint Experimentation, Transformation and Concepts Division. U.S. Department of Defense, Joint Staff J7.

Jürgen, S., 2003. Calculated Security? Mathematical Modelling of Conflict and Cooperation. In: Booß-Bavnbek, B., Høyrup, J., eds. Mathematics and War. Birkhäuser, Basel, 3rd edition, pp. 390-412.

Karlin, S., Taylor, M.H., 1975. A First Course in Stochastic Processes, 2nd edition, Academic Press.

Kee, J.E., 1999. At What Price? Benefit-Cost Analysis and Cost-Effectiveness Analysis in Program Evaluation. The Evaluation Exchange, Volume V, Number 2/3, Cambridge, MA.

Kerr, C., Phaal, R., Probert, D., 2006. A Framework for Strategic Military Capabilities in Defense Transformation. 11th International Command and Control Research and Technology Symposium (ICCRTS), Cambridge, MA.

Kim, B.S., 2012. Measuring Technological Change – Concepts, Methods, and Implications in Technological Change. In: Technological Change. Teixeira, A., ed., IntechOpen. https://doi.org//doi:10.5772/2314

Kim, D., Moon, H., Park, D., Shin, H., 2017. An Efficient Approximate Solution for Stochastic Lanchester Models. Journal of the Operational Research Society 68(11), pp. 1470-1481.

Kingman, J.F.C., 2002. Stochastic Aspects of Lanchester's Theory of Warfare. Journal of Applied Probability 39(3), pp. 455-465.

Kosola, J., 2013. Vaatimustenhallinnan opas (in Finnish). Sotatekniikan laitos, Julkaisusarja 5, 12 Maanpuolustuskorkeakoulu, Tampere.

Kress, M., 2020. Lanchester Models for Irregular Warfare. Mathematics 2020, 8(5)737. https://doi.org//doi:10.3390/math8050737

Kudryavtsev, D., Grigoriev, L., Bobrikov, S., 2014. Strategy-Focused and Value-Oriented Capabilities: Methodology for Linking Capabilities with Goals and Measures. The 1st International Workshop on Capability-oriented Business Informatics (CoBI 2014), Geneva, Switzerland, CEUR Workshop Proceedings, 1182:15-26.

Kuikka, V., 2018a. Influence Spreading Model Used to Analyse Social Networks and Detect Sub-communities. Computational Social Networks 5:12. https://doi.org/10.1186/s40649-018-0060-z

Kuikka, V., 2018b. Terrorist Network Analyzed with an Influence Spreading Model. In: Cornelius S., Coronges K., Concalves B., Vespignani A., eds. Complex Networks IX. Complenet 2018. Springer Proceedings in Complexity. Springer, Chan. https://doi.org/10.1007/978-3-319-73198-8_16

Kuikka, V., 2019a. Methods for Modeling Military Capabilities. IEEE Systems Conference (Syscon 2019), Orlando, FL.

Kuikka, V. 2019b. Modeling Network Resilience and Utility of Services. IEEE Systems Conference (Syscon 2019), Orlando, FL.

Kuikka, V., 2019c. A General Method for Detecting Community Structures in Complex Networks. In: Cherifi H., Gaito S., Mendes J., Moro E., Rocha L., eds. Complex Networks and their Applications VIII (Complex Networks 2019), Studies in Computational Intelligence vol. 881, Springer. https://doi.org/10.1007/978-3-030-36687-2_19

Kuikka, V., 2021. Modelling Community Structure and Temporal Spreading on Complex Networks. Computational Social Networks 8:13. https://doi.org/10.1186/s40649-021-00094-z

Kuikka, V., Nikkarila, J.-P., Suojanen, M., 2015. A Technology Forecasting Method for Capabilities of a System of Systems, Portland International Conference on Management of Engineering & Technology (PICMET 2015), Portland, Oregon, USA.

Kuikka, V., Syrjänen, M., 2019. Modelling Utility of Networked Services in Military Environments. International Conference on Military Communications and Information Systems (ICMCIS 2019), Budva, Montenegro.

Kuikka, V., Nikkarila, J.-P., 2019. Cooperation and Leadership in Military Organisations Modelled with Methods of Social Network Analysis. International Conference on Military Communications and Information Systems (ICMCIS 2019), Budva, Montenegro.

Kuikka, V., Latikka J., 2018. Mathematical Method for Selecting Team Members From a Social Network. IJERD, (14)2/2, pp. 1-15

Kuikka, V., Peltotalo, S., 2020. Tietoliikenneverkkojen ja Tietoteknisten Palveluiden Tarjoaman Sotilaallisen Suorituskyvyn Matemaattinen Mallintaminen (in Finnish). Tiede ja Ase 2020.

Lanchester, F.W., 1914. Aircraft in Warfare: The Dawn of the Forth Arm – No. V., The Principle of Concentration. Engineering (98), pp. 422-423.

Lappi, E., Pakkanen, M., Åkesson, B., 2012. An Approximative Method of Simulating a Duel. Proceedings of the Winter Simulation Conference (WSC'12), Berlin, Germany, pp. 1-10.

Lauren, M.K., 2002. A Fractal-Based Approach to Equations of Attrition. Military Operations Research 7(3), pp. 17-29.

Levitin, G., Finkelstein, M., Dai, Y., 2020. Optimal Preventive Replacement Policy for Homogeneous Cold Standby Systems with Reusable Elements. Reliability Engineering & System Safety, vol. 204, 107135.

Lindquist, J.M., Teutsch, C., Torres, E., 2017. A New Use for an Old Tool: Markov Chains and the Afghanistan Warfighter. Military Operations Research 22(2), pp. 5-18.

Ling, M.L., 2005. Proposed Network Centric Warfare Metric: From Connectivity to the OODA Cycle. Military Operations Research (10)1.

Liu, Q., Xing, L., Zhou, C., 2019. Probabilistic Modeling and Analysis of Sequential Cyber-attacks. Engineering Reports 1:e12065, https://doi.org/10.1002/eng2.12065

Liu, X., Li, D., Ma, M., Szymanski, B.K., Stanley, H.E., Gao, J., 2020. Network Resilience. Accessed 16.4.2021 https://arxiv.org/abs/2007.14464

Low, L.J., 1995. Anatomy of a Combat Model. Report, Accessed 16.4.2021 http://www.militaryconflict.org/Anatomy%20of%20a%20Combat%20Model_1.pdf

Lucas, T.W., Dinges, J.A., 2004. The Effect of Battle Circumstances of Fitting Lanchester Equations to the Battle of Kursk. Military Operations Research 9(2), pp. 17-30.

Luss, H., 1999. On Equitable Resource Allocation Problems: A Lexicographic Minimax Approach. Operations Research 47(3).

MacKay, N., 2015. When Lanchester met Richardson, the Outcome Was Stalemate: A Parable for Mathematical Models of Insurgence. Journal of the Operational Research Society (66), pp. 191-201.

Martin, T., Zhang, X., Newman, M.E.J., 2014. Localization and Centrality in Networks. Physical Review E 90, 052808. https://doi.org/10.1103/PhysRevE.90.052808

Martino, J.P., 1993. A Comparison of Two Composite Measures of Technology. Technological Forecasting and Social Change (44), pp. 147-159.

McChrystal, S., 2015. Team of Teams: New Rules of Engagement for a Complex World, Portfolio/Penguin.

Moffat, J., Smith, J., Witty, S., 2006. Emergent Behavior: Theory and Experimentation Using the MANA Model. Journal of Applied Mathematics and Decision Sciences 054846, pp. 1-13. https://doi.org/10.1155/JAMDS/2006/54846

Moore, S.C., Stockfish, J.A., Goldberg, M.S., Holroyd, S.M., Hildebrandt, G.G., 1991. Measuring Military Readiness and Sustainability, National Defense Institute RAND report. Accessed 16.4.2021 R-3842-DAG

Morse, P.M., Kimball, G.E., 1951. Methods of Operations Research. M.I.T. Press, Cambridge, MA.

National Research Council, 2006. Defense Modeling, Simulation, and Analysis: Meeting the Challenge. The National Academies Press, Washington, DC. https://doi.org/10.17226/11726

NATO, 2019. NATO Glossary of Terms and Definitions. AAP-06 Edition 2019.

Neaga, E.I., Henshaw, M., Yue, Y., 2009. The Influence of the Concept of Capability-based Management on the Development of the Systems Engineering Discipline. Proceedings of the 7th Annual Conference on Systems Engineering Research, Loughborough University, UK.

Newman, M.E.J., 2010. Networks, An Introduction. Oxford University Press, Oxford.

Newman, M.E.J., 2012. The Mathematics of Networks. Center for the Study of Complex Systems, University of Michigan, Ann Arbor. Accessed 16.4.2021 http://www-personal.umich.edu/~mejn/papers/palgrave.pdf

Nikkarila, J.-P., Kuikka, V., 2020. Complex Network Analysis for Evaluation Effects of Asymmetry Established by Closed National Networks. IEEE International Conference on Cyber Warfare and Security (ICCWS), Norfolk, VA.

Ogryczak, W., Luss, H., Pióro, M., Nace, D., Tomaszewski, A., 2014. Review Article Fair Optimization and Networks: A Survey. Journal of Applied Mathematics, Hindawi Publishing Corporation, pp. 1-26.

Osipov, M., 1915. The Influence of the Numerical Strength of Engaged Forces on their Casualties. Translated by Helmbold, R.L, Rehm, A.S. In: Warfare Modeling, Military Operations Research Society, John Wiley & Sons, 1995, pp. 290–343.

Parnell, G.S., 2017. Trade-off Analytics, Creating and Exploring the System Tradespace. John Wiley & Sons., Hoboken, New Jersey.

Reese, T.F., 1971. A Finite Markov Chain Model of the Combat Process. Master's thesis, United States Military Academy. Accessed 16.4.2021 AD736113

Perry, W., 2003. Modeling Knowledge in Combat Models. Military Operations Research 8(1), pp. 43-55.

Saaty, T.L., 1990. The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. ISBN 0-07-054371-2, RWS Publications McGraw-Hill.

Sage, A.P., Cuppan, C.D., 2001. On the Systems Engineering and Management of Systems of Systems and Federations of Systems. Information Knowledge Systems Management 2, pp. 325-345.

Salim, A.S.R., Hamid, W.M., 2001. A Bayesian Stochastic Formulation of Lanchester Combat Theory. Military Operations Research 6(3), pp. 69-76.

Saperstein, A.M., 1994. Mathematical Modeling of the Effects of 'Capability' and 'Intent' on the Stability of a Competitive International System. Synthese 100(3), pp. 359-378. https://doi.org/10.1007/BF01063908

Scala, N.M., Howard II, J.P., eds. 2020. Handbook of Military and Defense Operations Research. Chapman & Hall/CRC Series in Operations Research.

Smith, C.J., Oosthuizen, R., 2011. Applying Systems Engineering Principles towards Developing Defence Capabilities. Presented at the Industrial, Systems and Engineering Management Conference, Stellenbosch ISEM 2011 Proceedings, South Africa, p. 103.

Smith, C.J., Oosthuizen, R., Harris, H., Venter, J.P.,Combrink, C., Roodt, J.H.S., 2012. System of Systems Engineering – The Link Between Operational Needs and System Requirements. South African Journal of Industrial Engineering 23(2), pp. 47-60.

Sozen, C., Basim K., Hazir K., 2009. Social Network Analysis in Organizational Studies. International Journal of Business and Management (1)1.

Speight, L.R., 2001. Lanchester's Equations and the Structure of the Operational Campaign: Within-Campaign Effects. Military Operations Research 6(1), pp. 81-103.

Speight, L.R., 2002. Lanchester's Equations and the Structure of the Operational Campaign: Between-Campaign Effects. Military Operations Research 7(2), pp. 15-43.

Suojanen, M., Kuikka, V., Siivonen, P., Nikkarila, J.-P., Hämäläinen, J., Laari, T., Mattila, M., Rusila, T., Teittinen, K., Vestama, T., 2014. Hierarchic Technology Forecasting Model – What Can Be Data Farmed? In: Meyer, T., Horne, G.E., eds. Scythe: Proceedings and Bulletin of the International Data Farming Community, Issue 15, Workshop 27, pp. 30-33.

Suojanen, M., Kuikka, V., Nikkarila, J.-P., Nurmi, J., 2015. An Example of Scenario-based Evaluation of Military Capability Areas – An Impact Assessment of Alternative Systems on Operations. IEEE International Systems Conference (Syscon), Vancouver, Canada.

Taylor, J.G., Neta, B., 2001. Explicit Analytical Expression for a Lanchesterian Attrition Rate Coefficient for Bonder and Farrell's m-Period Target Engagement Policy. Working Paper #5. DTRA Project.

Tellis, A.J., Bially, J., Layne, C., McPherson, M., 2000. Chapter 7 In: Measuring National Power in the Postindustrial Age, RAND Monograph. Accessed 16.4.2021 https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1110.1.pdf

Tugrul, M.A., Daim, U., Jetter, A., 2013. A Review of Scenario Planning. Futures (46), pp. 23-40.

Turkes, T., 2000. Fitting Lanchester and Other Equations to the Battle of Kursk Data. Master's thesis. Department of Operations Research, Naval Postgraduate School, Monterey, California.

Tzeng, G.-H., Shen, K.-Y., 2017. New Concepts and Trends of Hybrid Multiple CriteriaDecision Making. CRC Press, Boca Raton, FL.

Walk, S.R., 2012. Quantitative Technology Forecasting Techniques. In: Teixeira, A., ed. Technological Change, InTech.

Washburn, A., Kress, M., 2009. Combat Modeling; Springer: New York, NY, USA.

Webb, M., 2008. Capabilities-based Engineering Analysis (CBEA). In: Minai, A., Braha, D., Bar-Yam, Y., eds. Unifying Themes in Complex Systems, Proceedings of the Sixth International Conference on Complex Systems, Springer Berlin Heidelberg, pp. 540-547.

Willard, D., 1962. Lanchester as Force in History: An Analysis of Land Battles of the Years 1618 – 1905. No AD297375, DTIC, Alexandria, Virginia.

Windrum, P., Diaz, C., Filiou, D., 2009. Exploring the Relationship between Technical and Service Characteristics. Journal of Evolutionary Economics (19), pp. 567-588.

I

Modelling the Impact of Technologies and Systems on Military Capabilities

Vesa Kuikka and Marko Suojanen

Journal of Battlefield Technology, 17(2), 2014, pp. 9-16

MODELLING THE IMPACT OF TECHNOLOGIES AND SYSTEMS ON MILITARY CAPABILITIES

Vesa Kuikka[1] and Marko Suojanen[1]

Abstract: Capabilities are commonly used in planning and modelling of military power. In this work we introduce a method for modelling the impact of future technologies and new systems on capabilities. Based on probability theory, a mathematically tractable definition of capability is given. The model relates the capability of a system to the capability area (such as situational awareness, for example). We show how to "drill down" from high level capabilities into system capabilities. Relationships between different parts of the system of systems are defined in a quantified manner. The modelling method is demonstrated with data from a questionnaire where changes of capabilities were evaluated.

INTRODUCTION

Military capabilities have been characterized by qualitative and quantitative terms. Measuring military capability and the difficulty of quantifying military capability in a single, definite measure has been discussed over the years [1-6]. In this paper we present a probabilistic model and propose a quantitative definition of capability.

The concept of capability has been used to express the level of will, amount of troops and armament. Combat capability is a generalized characteristic of the quantity and quality of forces and assets [7,8]. In the US the Joint Capability Areas (JCA) is a standardized set of definitions that cover the complete range of military activities [9]. Comparable standardization work has been completed in many countries [10,11]. The concept of capability can be extended outside the military context.

Capability is an abstract concept, not directly applicable to mathematical calculations. Here, we define the capability as the probability of a successful mission. The system capability measure is defined as the probability of successful system operation in a mission. The definition gives us the possibility to set numeric values to the capabilities. The definition is confined to the mission, later we use the term scenario, which gives all the necessary information needed for the evaluation of the probability value. For example, the time frame available for the mission has an effect on the probability of mission success. The probability of mission success as a measure of capability has its limitations. A single number does not describe details of assets and forces.

A problem in capability modelling is how to determine the relationship between the individual system capabilities and the higher level capability areas. Many attempts have been made by heuristic models where relations are described by weighting factors. Unclear definitions of terms and concepts can make the interpretation of the results difficult or infeasible. The relations between different capabilities are taken into account with the help of probabilistic relations. These definitions quantify the concept of capability and make the modelling possible.

Standardization of systems architecture framework is going on in NATO and several other organizations. Architecture frameworks offer us the connection between capabilities, systems and technologies. The level of detail in the model is left to the modeller to decide on the grounds of the requirements and the available input data. System of systems engineering is an integral method in the modelling [5,6]. One proposed definition of system of systems is "Modern systems that comprise system of systems are not monolithic, rather they have five common characteristics: operational independence of the individual systems, managerial independence of the systems, geographical distribution, emergent behavior and evolutionary development." [12,13]

We present a generic method that can be extended according to the specific needs of the modelling. The method allows modelling a limited set of systems and technologies. It is possible to use the method to high level capabilities when only one level of hierarchy below the main capability areas is considered. This kind of modelling can provide a tool for understanding the "big picture" of the capability structure. The high level model can further be extended to lower levels of system hierarchy and desired details of the model.

Not the entire defense system and its complexities need to be included. Trying to model a large complex system can lead to a black box where dependencies between the systems are not tractable. In our method the systems are modelled as separate building blocks. The problem is divided into manageable entities where the modelling is encapsulated in the system of systems philosophy. The dependencies between systems and components are modelled with well-known probabilistic concepts without ad hoc parameters.

Quantifying the concept of capability enables us to make a novel approach to the problem. It may be easier for experts to evaluate the overall capability change of capability areas induced by a new system or technology than to evaluate the capability change of the individual system itself [14]. There is a side effect that in addition to the technological change also tactical or other improvements in the environment may occur at the time. At first glance, this seems to be a drawback, but after all, the military is interested in the capability areas values, to pinpoint the original cause is of secondary interest to the officers. Because of the interrelationships, evaluation of the individual systems capabilities alone, without taking into account other changes, is difficult or even impossible.

PRINCIPLES OF THE MODEL

Our study is not restricted by a particular specification of capability hierarchy or even the concept of capability. We can

1. Finnish Defence Research Agency, PO Box 10, FI- 11311, RIIHIMÄKI, Finland, Tykkikentäntie 1.

ISSN 1440-5113

map the effects of system development to any set of entities. For example, alternative entities for modelling military readiness and feasibility have been presented in [13].

In our study, we conducted a questionnaire where changes in the most important capability areas were evaluated as a result of deploying a new system or a set of systems. The following capability areas have been selected for the study: protection, awareness and engagement. Because we have a quantitative model, we can calculate the changes of individual system capabilities by reverse-engineering. The system capability values include also other changes in the environment, e.g. better tactics or procedures. These may or may not be directly connected with the deployment of new systems or technologies. As we said before, it is difficult to separate the causes. If desired, the method still has the option to model tactics as a separate virtual system.

Different combinations of systems and disposal of old systems can be calculated with the model. Alternative defense systems can be compared to maximize the capability and to minimize the costs. The common situation is that several systems contribute to the same capabilities. In this way, the systems are alternatives to perform the same mission functionalities. In military conflicts to maximize the impact, or the probability of success, the systems are applied as complementary resources of capability.

The systems are used in parallel to back up each other. When the systems are both functioning, one of the two systems gives no additional capability. From probability theory the formula for the probability of union of events, that are not necessarily mutually exclusive, is used in the model.

In addition, our goal is to forecast the effects of technology changes on capabilities in time scale when only the development of the technologies is known. This is possible because we have the relationship of technologies to system capabilities and the relationship of system capabilities to capability areas. The reliability of the calculation depends on the level of details in the model and the accuracy of the input parameters.

In this work we combine capabilities, architectures and system of systems ideas with mathematical probability models [15,16]. Mathematical ideas of reliability and graph theory [17] can be utilized in the modelling.

MATHEMATICAL MODELLING OF MILITARY CAPABILITIES

We use basic probability theory and ideas of system of systems for constructing the model. According to the definition of system of systems cited earlier in this paper, the capability areas can be modelled separately. We apply the model for three capability areas: protection, awareness and engagement and two systems: satellites (SAT) and unmanned aerial vehicles (UAV). In the questionnaire the initial capability values were given for the three scenarios. The capabilities, i.e. the probabilities of mission success, were evaluated for 1 year and 10 years of technological development by the respondents.

Despite the fact that the capabilities can be modelled separately, this does not preclude that they can depend on the same variables while at the same time the capabilities may be statistically independent. In fact, the simplest assumption is that the total capability is the product of the capability areas values. In other words, they are considered as serial systems. On the lower hierarchical levels, the systems may be in series or in parallel. In the most natural arrangement the satellites, the UAVs and the parallel systems are in parallel. Each one of these systems performs the same functionalities on a capability area but with different capability values. In this paper we call the system of systems *k* in Figure 1 the parallel system. The parallel system can be a land, naval or airborne system or a combination.

We know that satellites and UAVs have functionalities that are not in parallel, for example satellites do not have the firing capability against ground objects. However, satellites can create considerable engagement capability by providing target acquisition. We have two points here. Target acquisition may be defined as a part of awareness or engagement capability area. The total capability value does not depend on the specification. The second point is that better awareness enables the engagement and the protection functionalities which induce additional capability improvement. The capability areas are related in a complex manner. In evaluations and calculations the capabilities of functionalities must be taken into account once, and only once. If the functionalities in the capability areas specification do not overlap and cover the whole range, no problems in assigning the capability values correctly will occur.

In our model we have the satellite system, the UAV system, the parallel system and the auxiliary system in the scenario. The parallel system has, by definition, the same functionalities as the satellite and the UAV systems. The auxiliary system has the functionalities of the capability area which are not included in the parallel system of systems of the satellites, the UAVs and the parallel system in the scenario. We denote the system capabilities of the satellites, the UAVs, the parallel system and the auxiliary system X_1, X_2, X_k, and X_m, correspondingly (Figure 1).

The probability values P_1, P_2, and P_{12} for the satellites, the UAVs and the combination correspondingly are from the questionnaire data. The initial probability values P_0 are known. The probability values X_1, X_2, X_k, and X_m are calculated from the model in Figure 1. The model provides the connection between capability areas and systems, i.e. the mapping from P_0, P_1, P_2, and P_{12} to X_1, X_2, X_k, and X_m.

From basic probability theory we get the following equations:

$$X_m X_k = P_0 \tag{1}$$

$$X_m(X_1 + X_k - X_1 X_k) = P_1 \tag{2}$$

$$X_m(X_2 + X_k - X_2 X_k) = P_2 \tag{3}$$

$$X_m(X_1 + X_2 + X_k - X_1 X_2 - X_1 X_k - X_2 X_k + X_1 X_2 X_k) = P_{12} \tag{4}$$

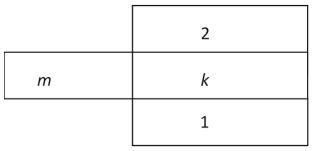

Figure 1. System of systems of the study, 1 = satellite system, 2 = UAV system, *k* = parallel system and *m* = auxiliary system. Systems 1, 2 and *k* are in parallel and system *m* is in series with the system of 1, 2 and *k*.

In the first equation no satellites and no UAVs are available. In the second and third equations the satellites or the UAVs are in use and in the last equation both the satellites and the UAVs are in use. The initial capability P_0 consists of X_k and X_m. The graph in Figure 1 illustrates the arrangement.

Equations (1–4) represent one capability area in one scenario at time *T*. In Equations (1–4) the variables P_0, P_1, P_2, and P_{12} are known and the variables X_1, X_2, X_k, and X_m are unknown. Expressing the capabilities with the help of the initial capability $P_0 = p_0$ and the capability changes p_1, p_2, and p_{12} we have, $P_1 = p_0 + p_1$, $P_2 = p_0 + p_2$ and $P_{12} = p_0 + p_{12}$. From Equations (1–4) we get:

$$X_1 = \frac{p_1+p_2-p_{12}}{p_2} \tag{5}$$

$$X_2 = \frac{p_1+p_2-p_{12}}{p_1} \tag{6}$$

$$X_m = p_0 + \frac{p_1 p_2}{p_1+p_2-p_{12}} \tag{7}$$

$$X_k = \left(1 + \frac{p_1 p_2}{p_0(p_1+p_2-p_{12})}\right)^{-1} \tag{8}$$

Equations (5) and (6) do not depend on p_0 which is a feature of the model in Figure 1. It is intuitive that the system capabilities do not depend on the outside environment of the systems. The idea of Equations (5–8) is to factor out the dependency on systems and on scenarios more explicitly. This makes the model extendable with combining additional systems with the satellite and the UAV systems. New systems are inserted in parallel or in series depending on the system functionalities.

Another consequence is that it is not necessary to investigate every pair of large number of systems. It is enough to find out all the system capability values X_i of the model. The values can be evaluated from experts' views or by simulation and modelling.

The value p_0 depends a great deal of the scenario while the values p_1, p_2, and p_{12} depend less on the scenario. This can be seen from Table 1 or Figures 2a-c. The initial value p_0 describes different positions of the opposing sides in the scenario. The capability changes p_1, p_2, and p_{12} describe the system capability changes in the scenario on the capability area level. The capabilities X_k and X_m. are dependent on p_0 because they constitute the initial position of the scenario as stated in Equation (1).

Equations (5-8) are used separately for all the cases in our study: three scenarios and four time frames. The same form of the equations is used for the protection, the awareness and the engagement capability areas. The total capability value is calculated from the equation:

$$P = P_{Prot} P_{Awa} P_{Eng} \tag{9}$$

The results from Equation (9) for the scenarios, systems and time frames are shown in Table 2. The combination of the satellites and the UAVs is calculated from Equation (9) with the values of P_{12}.

Equation (9) follows from the assumption of statistical independence. This is a consequence of the fact that the capability areas are defined according to the normalization principles of the traditional data modelling practice where no overlapping of concepts is allowed. The same idea is inherited from the definitions of the capability areas to the system level functionalities. The assumption of statistical independency and the interdependency of capabilities are two different and coexistent properties of our modeling problem. In our approach the exact form of the dependence between the capability values or between the system capability values is not needed. Defining the necessary variables and the subsequent modelling would be difficult and dependency on tactics and other non-materiel conditions makes the task even harder.

The capability values X_1, X_2, and X_k of the three systems are solved assuming that they are operating in parallel. Corresponding procedures work for different combinations of serial and parallel systems. The method allows different serial - parallel models for every capability area. Different models follow from different functionalities of a system in the capability areas. In our study, the arrangement in Figure 1 is assumed to be the best first degree modelling structure for the three capability areas.

THE QUESTIONNAIRE DATA

We demonstrate the method with an example. The questionnaire among ten people was performed in the International What If? workshop in January 2014 [18]. Half of the group were officers and the other half civilian. None of them were expert in the new systems under consideration. Scarce background information was given and there was very limited time to prepare for the questionnaire. The numerical results may not be valid for decision-making purposes. However, the data is sufficient for studying the modelling problem.

Technological development of two systems was evaluated in three scenarios for immediate deployment and within a time frame of ten years. For the immediate deployment, one year was allowed for procurement, training and planning. The system candidates were the satellites and the UAVs. The combined use of the satellites and the UAVs was evaluated in the same way. The number and the exact type of satellites and UAVs were not given beforehand and the respondents had to specify these facts eligible for the scenarios. The capability estimations for three capability areas were asked.

In the scenarios these three capability areas are assumed to be the most important, but for the modelling of complete defense system capability, all the capability areas should be included in the model.

The respondents had no information about the model of Figure 1 or Equations (1–8). Nevertheless the respondents evaluated the capability values of the defense system that included other systems and tactics of the scenarios.

In the beginning ($T = 0$) in Scenario 1 the two sides have equal capabilities, in Scenario 2 the adversary has superior capacity and in Scenario 3 is the same as Scenario 2 but from civilian authorities' point of view. In the questionnaire, effects of UAVs and satellites will be analyzed from the BLUE perspective. BLUE side does not have any satellite or UAV assets but, as in Scenarios 2 and 3 for example, the RED nation has both satellites and even UAVs with attack capability. In the first scenario the main mission of the BLUE side is to take control over the RED island in order to build radar and missile launch systems to the island. There is a number of sea-, air- and land-based capabilities available for both BLUE and RED nations. In Scenarios 2 and 3, RED as a military superpower has conducted limited air-operations against BLUE air defense and harbors. Missions of the BLUE side are to defend areal integrity, to protect critical infrastructure and population and to keep sea lines open to ensure trade traffic. RED has control over its area and partially the BLUE's airspace. In Scenario 3, BLUE nation is responsible of the security of civilians in spite of RED nation's hostile actions. The three scenarios have been described in more detail in [15,19].

In the questionnaires individual opinions affect the values given for the probability values. An objective approach would be calculating the corresponding probabilities by using simulation tools or by applying other mathematical models. On the other hand, the results can be taken as describing subjective probabilities. We omit the human behaviour considerations in the questionnaire and keep in mind the limitations of interpreting the results strictly as the probability of mission success.

In Table 1 the average capability changes in the questionnaire are shown. Note that the changes are asked at the level of capability areas, not for the separate satellite or the UAV systems. The initial values for the protection capabilities are 0.7, 0.5, and 0.5 for the Scenarios 1–3. The initial values for the awareness capabilities are 0.8, 0.6, and 0.4 and the initial values for the engagement capabilities are 0.9, 0.3, and 0.4. As an example in Scenario 1 after one year the awareness capability is 0.8 + 0.072 = 0.872 (87.2 %) when the satellites and the UAVs are in use.

All the results of this paper can be calculated from Equations (5-8) with the information given in Table 1 and the initial values.

In the model we have 57 parameters: p_i, and $p_{i,T,C,S}$ where i={1, 2, 12}={SAT, UAV, SAT and UAV}, T={1 year, 10 years}, C={Protection, Awareness, Engagement} and S={Scenario 1, Scenario 2, Scenario 3}. With 57 parameters and 57 unknown variables we have merely changed variables so the model and the data match perfectly. The new variables X_1, X_2, X_k, and X_m describe the capabilities of individual systems i, not the capabilities $P_{i,T,C,S}$ of the entire system of systems.

In the second part of the questionnaire the effects of the development within 10-year and 20-year timeframes in technology areas on a set of functionalities were asked. A list of five functionalities was used: surveillance, communication, engagement, logistics and deception. The functionalities considered are connected to specific capability areas. A selected list of seven technology areas was used: sensor technology, material technology, communications technology, stealth technology, energy source technology, manufacturing technology and autonomous operation technology. For our modelling purposes we assume that they represent the general trend of technology development in 10 and 20 years. These data are used to extrapolate the capability changes from 10 years to 20 years. All the results for 20 years in this paper are from the extrapolation calculation.

In Table 2 the average values of total capability values after 0, 1, 10, and 20 years are given. The values for 0 year have been given beforehand to the respondents. The values after 1, 10, and 20 years have been calculated from Equation (9). The values after 1 and 10 years have been calculated from the questionnaire data and the values after 20 years from the extrapolation.

		1 year			10 years		
		Sce1	**Sce2**	**Sce3**	**Sce1**	**Sce2**	**Sce3**
Prot	**SAT**	0.030	0.048	0.034	0.055	0.082	0.077
	UAV	0.042	0.035	0.043	0.076	0.100	0.085
	both	0.064	0.077	0.053	0.121	0.153	0.097
Awa	**SAT**	0.033	0.040	0.071	0.063	0.083	0.157
	UAV	0.061	0.068	0.089	0.085	0.123	0.173
	both	0.072	0.079	0.121	0.104	0.151	0.208
Eng	**SAT**	0.002	0.033	0.018	0.016	0.096	0.036
	UAV	0.006	0.060	0.044	0.024	0.168	0.110
	both	0.007	0.062	0.052	0.025	0.193	0.126

Table 1. The average changes of the capability area values from the questionnaire.

	SAT			UAV			SAT & UAV		
	Sce1	**Sce2**	**Sce3**	**Sce1**	**Sce2**	**Sce3**	**Sce1**	**Sce2**	**Sce3**
0 y	0.50	0.09	0.08	0.50	0.09	0.08	0.50	0.09	0.08
1 y	0.55	0.12	0.11	0.58	0.13	0.12	0.60	0.14	0.13
10 y	0.60	0.16	0.14	0.63	0.20	0.17	0.69	0.24	0.19
20 y	0.62	0.19	0.16	0.67	0.26	0.21	0.74	0.32	0.24

Table 2. The average values of the total capability values after 0, 1, 10 and 20 years calculated from the questionnaire with Equation (9).

ANALYSIS OF THE DATA

The analysis in this study is confined to modeling issues, the military arguments of capability changes are only shortly described in this paper. Analysis of the qualitative aspects and textual comments given by the respondents of the questionnaire was conducted in [19].

Next, we try to find the most significant features of the data. We use an elementary method where we compare the values $p_{i,T,C,S}$ with the average values $p_{i,T,S}$, $p_{i,T,C}$, and $p_{T,C,S}$ where the average is taken over indexes *C, S* and *i* correspondingly. We say that if the capability value $p_{i,T,C,S}$ is 5% or more higher than the average value $p_{i,T,S}$, the capability *C* with the indexes *i, T, S* has increased significantly. We make the following observations: In Scenario 3 awareness is 7–13% more than the average after one and ten years and in Scenario 2 engagement is 9% more than the average after 10 years with the UAVs. In Scenario 3 engagement is 11% less than the average after 10 years with the satellites and with both the satellites and the UAVs. We have proceeded in this manner and collected the results of the analysis in Table 3. Other differences are less than 5% which are inside the error margins of the data.

In the following, because of the high number of dimensions of our study we analyze only the awareness capability area to demonstrate the value of the model. In Figures 2a–c and 3a–d the awareness capability changes and the system awareness capabilities in Scenarios 1–3 are shown. The analysis of the protection and the engineering capability areas can be performed in the same manner.

The most outstanding observation in Table 3 is the improvement in the awareness capability in Scenario 3 because this occurs when the satellites, the UAVs and both systems are in use after one and ten years. The first conclusion may be that both the satellites and the UAVs are the origin of the increase of the awareness capability values. However, from Figures 3 a–d we detect that this conclusion is incomplete. After one year the system capability of the satellites and the UAVs is moderate and the actual reason for the improved awareness capability is the increased capability of the auxiliary system as can be seen from the high value $X_m - p_0$ in Figure 3d. We have subtracted the constant p_0 from X_m to highlight the awareness capability in Scenario 3.

In Scenario 3 after ten years the awareness capabilities of the satellites and the UAVs increase and first conclusion is more correct but even then the auxiliary system capability has increased. As discussed earlier in the text, the satellites and the UAVs may act as enablers of functionalities in other systems. From Figure 3a we see that between 10 and 20 years the satellites give no more capability in Scenario 3 which is not obvious from Figure 2a.

The textual answers given by the respondents of the questionnaire validate the results based on the mathematical model of Equations (5–8). Some respondents stated that UAVs are mostly operated by the military and UAVs are not available in civilian operations. This is in good accordance with the model in Figure 3b because in 1 year the lowest capability given by the UAVs is in Scenario 3. Further from Figure 3b the utilization of the UAVs is increasing after 10 and 20 years in Scenario 3. Based on the questionnaire answers, some respondents recognized increased use of mini-UAVs at all levels of organizations' within 10-year time frame in Scenario 3.

The value of capability given by the auxiliary system can be explained by traditional infrastructure and systems of situational awareness, such as telephones, e-mail and manned patrols. These are mostly ground or surface functionalities. Figure 3d shows that there is a good chance of developing and utilizing these systems in 10 and 20 years in Scenarios 2 and 3. Probably, in Figure 3c, the low capability values of the parallel system is due to the fact that manned air surveillance is not a good alternative for civilian operations. Satellites are not targets for enemy actions and they provide new capability especially for civilian use according to Figure 3a. After 10 years the satellites give no capability increase.

From line nine in Table 3 we see that in Scenario 1 the awareness capability is increasing less than average with the UAVs and with both the UAVs and the satellites between 1 and 10 years. We find the explanations from Figures 3a–d. The UAVs give no extra capability and the satellites and the auxiliary system give moderate capability increase. This changes are not obvious from Figures 2a–c. The textual answers confirm that the UAVs are not always available for civilian use as stated before.

These examples are enough to recognize the complex interrelations of the capabilities and the fact that it is difficult to find out the underlying reasons for the capability changes. Our model is a tool to decompose the system and delve into the structure of the defense system.

1	*C*	$p_{i=\{1,2,3\},T=\{1y,10y\},C=\text{Awareness},S=3}$	+7% … +13%
2	*C*	$p_{i=2,T=10y,C=\text{Engagement},S=2}$	+9%
3	*C*	$p_{i=\{1,3\},T=10y,C=\text{Engagement},S=3}$	-11%
4	*C*	$p_{i=\{2,3\},T=10y,C=\text{Protection},S=3}$	-6% … -8%
5	*C*	$p_{i=3,T=10y,C=\text{Engagement},S=1}$	-6%
6	*C*	$p_{i=3,T=1y,C=\{\text{Engagement}\},S=\{1,2,3\}}$	-4% … -7%
7	*S*	$p_{i=\{1,2,3\},T=1y,C=\text{Engagement},S=2}$	+13% … +17%
8	*S*	$p_{i=\{1,2,3\},T=10y,C=\text{Engagement},S=2}$	+5% … +7%
9	*S*	$p_{i=\{2,3\},T=10y,C=\{\text{Awareness,Engagement}\},S=1}$	-5% … -9%
10	*i*	$p_{i=3,T=\{1y,10y\},C=\text{Protection},S=2}$	+5%
11	*i*	$p_{i=1,T=10y,C=\text{Engagement},S=\{2,3\}}$	-11% … -12%
12	*i*	$p_{i=3,T=10y,C=\text{Engagement},S=2}$	+7%

Table 3. The main features of the questionnaire data. In the second column the averaging index is shown.

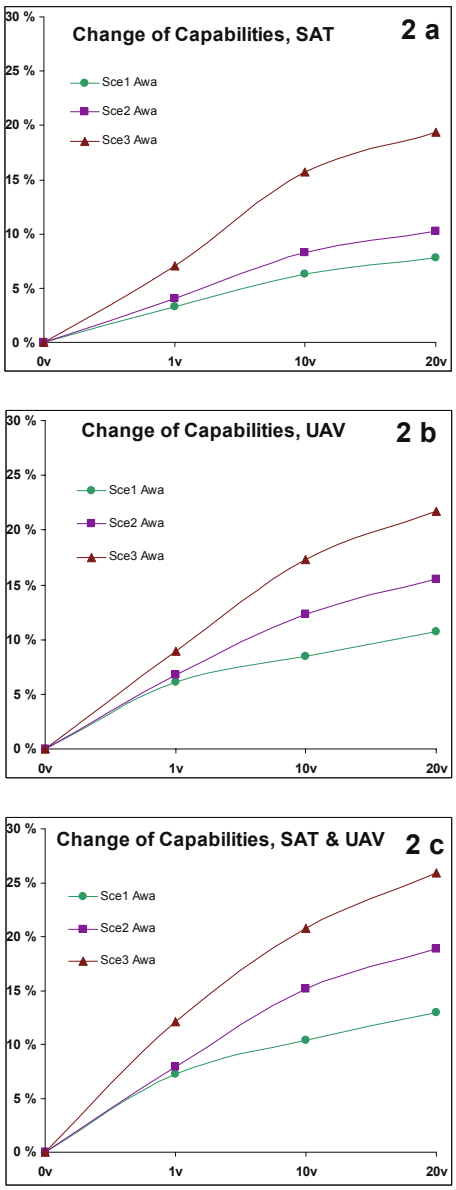

Figure 2. **The capability changes of awareness in Scenarios 1–3 when a) satellites b) UAVs and c) both satellites and UAVs are in use. The values after 0, 1, and 10 years are from the questionnaire data and the values after 20 years are the forecast results.**

On the other hand, we can start the analysis from Figures 3a–d. From Figure 3d we find that the auxiliary system capability value increases more in Scenario 2 than in Scenarios 1 and 3. In Scenario 2 the satellites, the UAVs and the parallel system give no extra capability. This can not be seen from Figures 2a –c.

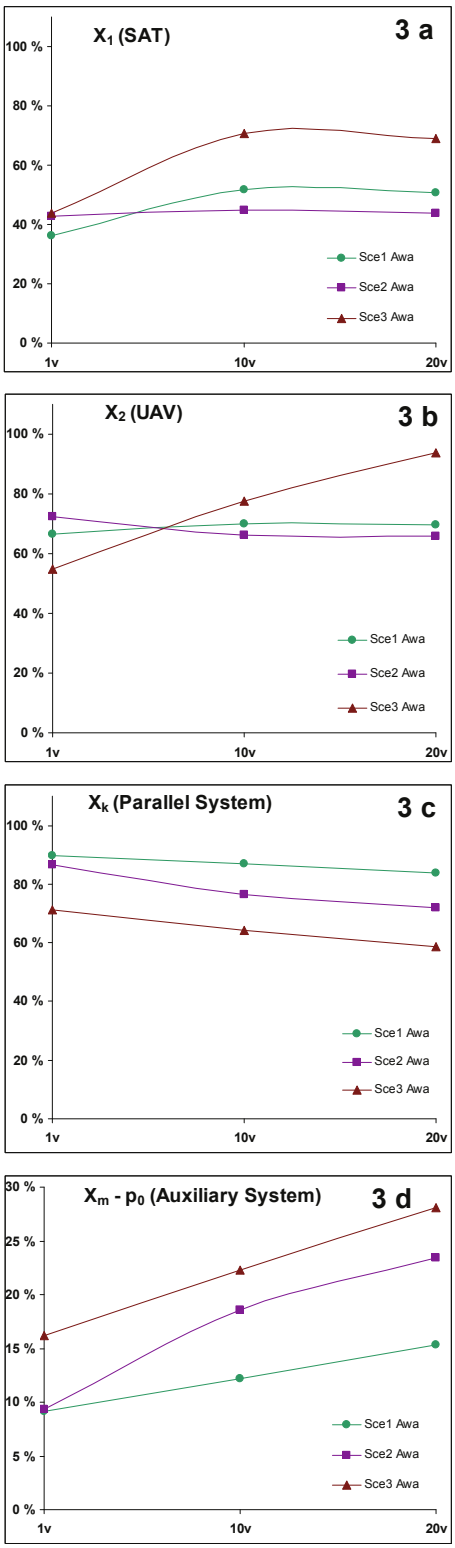

Figure 3. **The system capabilities of awareness in Scenarios 1–3 for X_1, X_2, X_k, and $X_m - p_0$. The values are calculated from Equations (5-8).**

Not all the explanations are found directly from the textual answers of the questionnaire. To demonstrate the use of the model we suggest the following explanations. Using manned helicopters or aeroplane are possible alternatives for air surveillance. This can explain the high values of Scenario 1 in Figure 3c. Developing traditional ground or surface infrastructure is not giving significant capability in Scenario 1 according to Figure 3d.

In all Scenarios 1–3 the capability of the parallel system is decreasing. The reason is that the UAVs and the satellites replace the functions of the parallel system. This cannot be seen from Figures 2a–c.

High values of X_m or X_k suggest modelling the auxiliary and the other existing systems in the scenario in more detail to raise the overall coverage of the model. In Figure 1 and Equations (7-8) the system content of *m* and *k* was unspecified. After modelling a new system or several systems in *m* or *k*, the corresponding capability changes are replace by the modelled set of systems.

High values of system capabilities may be brought about by numerical variability of the data. Low values of denominators of Equations (5–8) cause high sensitivity of the system capability values. Equation (1) determines the balance between X_m and X_k. This may be a reason for the unusual Scenario 1 engagement capability values of X_m and X_k. More analysis of the qualitative information gathered in the questionnaire or more accurate numerical data—that is, more respondents, is necessary to make the final conclusions in these cases.

CONCLUSIONS

A generic method of modelling military capabilities with probabilistic terms was presented. Capability was defined quantitatively as the probability of mission success.

Complex relations between systems' capabilities and capability areas were described by the union of non-exclusive events of system pairs. The utilization of the rules of probability theory simplified the modelling process significantly. Systems were modelled only once in the system of systems, not with all the possible pairs of systems. For example, adding a new system to a model of the satellites and the UAVs could be calculated without the recalculation of the model parameters of the satellites and the UAVs.

An introduction of new systems to the model or removing systems from the model gives directly the effect on the capability areas and total capability. In this way the model is a tool for comparing and planning alternative development and acquisition projects.

Different specifications of the capability areas can be used with the method and the desired level of details can be included in the model. The idea of the parallel systems and the auxiliary system can be used to point out the need for next iteration of a more detailed model.

The method was successfully demonstrated with questionnaire data and two systems, the satellites and the UAVs in three scenarios and three time frames. In future work the model can be extended with more systems.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Juhani Hämäläinen, Dr. Juha-Pekka Nikkarila and Maj. Pasi Siivonen for their contribution to the work done in International What If? Workshop. The authors wish to thank Dr. Gary Horne, Dr. Juhani Hämäläinen and M.Sc. Niina Nissinen for the opportunity to work in International What If? Data Farming Workshop 27 promoting the area of technology forecasting in the military context and publishing the initial work in the Scythe Proceedings.

REFERENCES

[1] The U.S. General Accounting Office, *Measures of Military Capability: A Discussion of Their Merits, Limitations and Interrelationships*, 1985.

[2] M.S.C. Craig, J.A. Stockfish, M.G. Goldberg, S.M. Holroyd, and G.G. Hildebrandt, *Measuring Military Readiness and Sustainability*, National Defence Institute RAND, 1991.

[3] C. Burris, D. Gonzales, J. McEver, D. Signori, M. Smeltzer, and H. Schoenborn, "Quantitative Capability Delivery Increments: A Novel Approach for Assessing DoD Network Capability", *Proceedings of the AIAA InfoTech@Aerospace Conference, American Institute of Aeronautics and Astronautics*, April 2010.

[4] E.N. Urwin et al, "Scenario-based Design and Evaluation for Capability", *IEEE 5th International Conference on System of Systems Engineering (SoSE)*, 22–24 June 2010.

[5] P.T. Biltgen and D.N. Mavris, "Capability-Based Quantitative Technology Evaluation for Systems-of-Systems", IEEE 7th *International Conference on System of Systems Engineering*, 16–18 April 2007.

[6] P.T. Biltgen and D.N. Mavris, "A Methodology for Technology Evaluation and Capability Trade-off for Complex System Architectures", *25th International Congress of the Aeronautical Sciences*, 2006.

[7] C. Wang and V.M. Bier, "Quantifying Adversary Capabilities to Inform Defensive Resource Allocation", in preparation.

[8] R.J. Dunn, "Measuring Military Capabilities: An Essential Tool for Rebuilding American Military Strength" *Backgrounder* No 2911, 16 May 2014.

[9] Defence Technical Information Center, http://www.dtic.mil/futurejointwarfare/strategic/jca_framewok.xls.

[10] J. Kosola, *Vaatimustenhallinnan opas*, Julkaisusarja 5. No 12, Sotatekniikan laitos, 2013.

[11] A. Hing, *Australian Defence Preparedness: Principles, Problems and Prospects: Introducing Repertoire of Missions (ROMINS) a Practical Path to Australian Defence Preparedness*, Australian Defence Studies Centre, Canberra, 2000.

[12] A.P. Sage and C.D. Cuppan, "On the Systems Engineering and Management of Systems of Systems and Federations of Systems", *Information, Knowledge, Systems Management*, Vol. 2, No. 4, 2001.

[13] P. Malone and L. Wolfarth, "System-of-Systems: An Architectural Framework to Support Development Cost Strategies", *IEEE Aerospace Conference*, 2012.

[14] X. Wang, C. Du, and Z. Cao, "Probabilistic Inversion Techniques in Quatitative Risk Assessment for Power System Load Forecasting", Proceedings of the 2008 *IEEE International Conference on Information and Automation* June 20–23, 2008, Zhangjiajie, China.

[15] V.V. Polenin, "Probabilistic Models in Operation Planning", *Military Thought*, April 2004.

[16] E.J. Offutt, J.P. Kharoufeh, and R.F. Deckro, "Distorted Risk Measures with Application to Military Capability Shortfalls", *Military Operations Research*, Volume 11, Number 4, 2006.

[17] L. Zhang, B. Mitchell, L. Falzon, M. Davies ,L.M. Kristensen and J. Billington, "Model-based Operational Planning Using Coloured Petri Nets", *6th International Command and Control Research and Technology Symposium*, June 2001.

[18] M. Suojanen et al, "Team 4: Hierarchic Technology Forecasting Model—What Can Be Data Farmed?", *Scythe Proceedings and Bulletin of the International Data Farming Community*, Issue 15 – Workshop 27, eds T. Meyer and G. Horne, pp. 30–33, May 2014.

[19] M. Suojanen, V. Kuikka, J-P. Nikkarila, and J. Nurmi, "An Example of Scenario-based Evaluation of Military Capability Areas—An Impact Assessment of Alternative Systems on Operations", proposal submitted to *Military Communications Conference*, MILCOM 2014, 2014.

Vesa Kuikka obtained the degree of Licentiate in Philosophy in physics in 1986 from the University of Helsinki and in mathematics in 2004 from Åbo Akademi University. He works at the Finnish Defence Research Agency in the technology forecasting group. He can be contacted by email at vesa.kuikka@mil.fi.

Marko Suojanen obtained his M.Sc.(EE) in electronics from the Tampere University of Technology in 2002. Following a number of years as a research scientist at Technical Research Centre of Finland and as an R&D engineer in private companies, he joined the Finnish Defence Research Agency, where his interests include technology forecasting, electronic warfare, and C4 systems. He can be contacted by email at marko.suojanen@mil.fi.

II

Dependency of Military Capabilities on Technological Development

Vesa Kuikka, Juha-Pekka Nikkarila and Marko Suojanen

Journal of Military Studies, 6(2), 2015, pp. 29-58

Available at https://doi.org/10.1515/jms-2016-0170

Vesa Kuikka, PhLic, Finnish Defence Research Agency, vesa.kuikka@mil.fi
Juha-Pekka Nikkarila, PhD, Finnish Defence Research Agency, juha-pekka.nikkarila@mil.fi
Marko Suojanen, M.Sc.(EE), Finnish Defence Research Agency, marko.suojanen@mil.fi

DEPENDENCY OF MILITARY CAPABILITIES ON TECHNOLOGICAL DEVELOPMENT

Keywords

Modelling military capability, technological forecasting, interdependencies between technologies

Abstract

Our goal is to get better understanding of different kind of dependencies behind the high-level capability areas. The models are suitable for investigating present state capabilities or future developments of capabilities in the context of technology forecasting. Three levels are necessary for a model describing effects of technologies on military capabilities. These levels are capability areas, systems and technologies. The contribution of this paper is to present one possible model for interdependencies between technologies. Modelling interdependencies between technologies is the last building block in constructing a quantitative model for technological forecasting including necessary levels of abstraction. This study supplements our previous research and as a result we present a model for the whole process of capability modelling. As in our earlier studies, capability is defined as the probability of a successful task or operation or proper functioning of a system. In order to obtain numerical data to demonstrate our model, we conducted a questionnaire to a group of defence technology researchers where interdependencies between seven representative technologies were inquired. Because of a small number of participants in questionnaires and general uncertainties concerning subjective evaluations, only rough conclusions can be made from the numerical results.

Introduction

New technologies can provide new and more effective military capabilities. New technologies can present threat or opportunity and their future development is uncertain. The uncertainty that characterises technologies mean that the military cannot know which emerging technologies mature to have profound impacts, how long that maturing will take nor the technological trajectory. Most emerging

technologies represent incremental improvements and enhance the competencies of the military. This kind of technological development presents few challenges to the military, although their adoption into existing platforms can be difficult. In contrast, it is new technologies that are radical, degrade competence and create new sources of military advantage along dimensions not traditionally valued or poorly understood by the military that are the focus of attention and concern. An emerging technology that undermines existing training, equipment and doctrine will have more impact on the military than one that complements or enhances existing military competencies. (James, 2013)

Fundamental concepts of this paper are capabilities, systems and technologies. Each of these has several different descriptions. We provide some of the most common definitions of these three concepts. Capability has been defined as the ability to achieve a specified wartime objective (win a war or battle, destroy a target set). The concept of capability has been used to express the level of will, amount of troops and armament. In many countries capability areas have been standardized as a set of specifications that cover the complete range of military activities. A system is a set of interacting or interdependent component parts forming a complex/intricate whole. Systems share common characteristics including structure, behaviour and interconnectivity. Technology is the collection of techniques, skills, methods and processes used in the production of goods or services. It is the state of knowledge of how to combine resources to produce desired products, to solve problems, fulfil needs, or satisfy wants.

In this paper we examine interdependencies of technologies and impacts of technologies on military capabilities. In our earlier research we have examined impacts of systems on military capabilities (Kuikka & Suojanen, 2014) and impacts of single technologies on military capabilities (Kuikka et al., 2015). In the present work we complement the modelling by taking into account influences of multiple technologies on system capabilities and top-level capability areas. In this respect, our aim is two-fold: to provide the missing piece of the modelling and give the big picture of the modelling results of our earlier research and this paper. To this end, we provide also a literature review of articles that are related to concepts and models of our work. The review is not a comprehensive survey of research in technology forecasting or capability modelling. The review serves as an introduction to the research field and other research projects comparable to our work. The theoretical and methodological literature ground our present work and guide for further investigation.

In this study, a functional dependency of military capabilities on technical development is presented. Questionnaire data is used as an input for modelling relationships between operational tasks, systems, technologies and capability areas. In this context, capability is defined as the probability of successful task or

operation, or functioning of a system. The functional form gives an approximation for calculating effects of different technological developments on capabilities.

In literature, describing military capabilities has been examined for a long time. Not many quantitative models have been published. Another problem is the lack of a measure for capabilities. In an earlier study, we have introduced also an alternative measure for strategy (Hämäläinen & Nikkarila, 2015). We propose a model based on a probabilistic measure of capabilities. The same measure is used in all levels of the model: capability areas, systems, tasks and technology areas.

The main idea is to show how to evaluate interdependencies between technology areas. For example, material technology has a significant effect on sensor technology. Progress in the first area implies progress in the latter area. Notice that interdependencies are not necessarily symmetric; for example sensor technology depends significantly on material technology, while the dependency in the opposite direction is weaker. Seven technology areas are considered: sensor, material, communications, stealth, energy source, manufacturing, and autonomous operation technologies. Five different operational tasks are used in the evaluation.

In the literature, different mathematical methods have been used when multiple interdependent variables are affecting the forecast. Examples are principal component analysis (Windrum et al., 2009), simulation methods and time series methods (De Gooijer & Hyndman, 2006). Typically these methods need more data than is available in our study. Some forecasting methods try to identify the underlying factors that might influence the variable that is being forecasted.

Judgmental forecasting methods incorporate intuitive judgments, opinions and subjective probability estimates. This characterization is also valid in our study. In addition, we construct models on the bases of questionnaire results. In this paper, we model interdependency of technology areas and dependency between technology areas and task capabilities. This study is one piece in larger modelling effort trying to increase our understanding of military capabilities. (Kuikka & Suojanen, 2014, Suojanen et al., 2015 and Kuikka et al., 2015 and references therein)

The operational analysis methodologies used in this work include systems modelling, probabilistic mathematical methods, statistical evaluations, information modelling and the use of questionnaires as the input data.

Related Work

In this section, a review of papers related to concepts, methods or application areas used in this paper is presented. First, we review articles on general theory of

technological change, which focus on questions like why certain technological developments emerge instead of others, are there regularities in the process of generation of new technologies and in technological progress thereafter, and is there a regularity in the functional relationship between the vast number of economic, social, institutional, scientific factors which are likely to influence the innovation process (Dosi, 1982). After general theory of technological change, we bring up a few articles on technometrics. Technometrics is a discipline that measures and evaluates technological change with important policy implications. The main techniques of technometrics and their potential and methodological difficulties have been presented in (Coccia, 2005). Scenario planning is a method for design and evaluation of capabilities, describing rare events and technology forecasting. Three articles related with scenario planning are referenced. Articles assessing weapon system capabilities, comparison between jet fighter aircraft and decision-analytic approach to reliability-based design optimization are related to methods in this paper. Finally, we give references closely related to our research of interdependency of technologies in the context of technological innovations and capability planning. In the following, we review related research articles and indicate their connections with our research. Similarities as well as different approaches with our method are commented after the references in each section.

The long-run economic innovation has been analysed as the interplay between supply-side and demand-side processes. These are technology-push and demand-pull forces. On the supply-side three different interacting development processes have been identified: growing productive efficiency, the emergence of new sectors and the increasing quality and differentiation of existing products. The time path of economic development cannot be explained by taking into account a supply-side view alone. Without an adequate demand, development processes cannot be generated. The situation can be described as the co-evolution of demand, innovation and supply (Saviotti & Pyka, 2013). Demand influences the selection among competing paradigms and the course of the paradigm after its inception. For example, in the history of computing technology a distinction can be made between periods in which either demand or knowledge development was the dominant enabler of innovation (van den Ende & Dolsma, 2002). Niches in evolutionary theories have been investigated in explaining radical technical change (Schot & & Geels, 2007). Radical change or technological discontinuity is defined as the establishment of a new sociotechnical regime. Sociotechnical regimes carry and store rules for how to produce, use and regulate specific technologies. The difference between niches results from differentiating between whether niches are internal or external to the prevailing sociotechnical regime and whether rules for design are stable or unstable within the niche. Our model describes the course of technological development after its inception with a linear model together with secondary multiplicative linear effects on other technologies. The dependencies between technology areas are linear but the overall effects on capability values are non-linear due to the interdependencies between technology areas.

The model of technological paradigms and trajectories (Dosi, 1982) account for both continuous change and discontinuous change in technological innovations. Continuous changes are related to progress along a technical trajectory defined by a technical paradigm, while discontinuities are associated with the emergence of a new paradigm. Technological paradigm has been defined in accordance with a set of procedures, a definition of relevant problems and the specific knowledge related to their solution. Technology trajectory is defined as the direction of advance within a technology trajectory. A radical innovation is founded on the creation of a new set of technology solutions, and results in a new trajectory that is qualitative different (Dosi, 1982). A model of technological evolution based on replicator dynamics has been presented (Saviotti & Mani, 1995) where the relationship between variety and competition has been studied. Probabilistic entropy statistics and scaling trajectories in 143 civil aircraft designs have been analysed in terms of changes in the product characteristics (Frenken & Leydesdorff, 2000). Distinction has been made between technical and service characteristics. Technical characteristics were defined as variables that can directly be manipulated by producers. Variables that users take into account in their purchasing decisions were considered as service characteristics. In our model, service characteristics can be identified with system services or functionalities and technical characteristics can be identified with technical features or attributes. Our basic model is designed for linear technological changes in technology areas including secondary effects on other technology areas. However, the model can be modified to include jumps triggered at specified levels. This can be easily implemented in a spreadsheet application. The secondary effects follow automatically. Even a more general functional form could be used, for example, exponential development. Our model takes into account the first order secondary effects in other technology areas, but the model does not account for 'economics of scale' which might support an exponential functional form.

The complex relationship between technical and service characteristics has been explored in (Windrum et al., 2009). Principal component analysis (PCA) has been used as a method of the analysis of a dataset of mobile phone handsets. Technological trajectories by means of a detailed case-study of the evolution of tank technology between 1915 and 1945 has been analysed with principal component analysis in (Castaldi et al., 2009). A hierarchic conceptualisation of tank technology with technical and service characteristics has been used in the modelling. Our model of system of systems consists of serial and parallel systems and subsystems. Even a simple system structure may describe high level capabilities and their long term development as well as more detailed hierarchic models or purely statistical methods such as PCA.

The scale of innovative intensity (SIIN) based on the economic impact of the technological innovation on the economic system has been used as a theoretical

measure for technological change (Coccia, 2003). The SIIN is similar to the seismic scale of measuring the intensity of earthquakes. Three families of complexity models of technology innovation have been discussed in (Frenken, 2006): fitness landscape models, network models and percolation models. The models are capable of analysing complex interaction structures while avoiding over-parameterisation. Technological developments in the network connecting patents have been analysed in (Schoen et al., 2012). In our model, a probabilistic measure for high level capabilities and system capabilities is proposed. This is a quantitative and intuitive measure suitable for all levels of capability modelling. The probability of success in an operation is considered to be a natural selection in the military context. Especially, when modelling complex interactions and structures, it is important to have an understandable and quantitative method. High uncertainties in long term forecasting support a simple model instead of more sophisticated considerations.

In technology planning, forecasting, strategic analysis, foresight studies, scenarios are used to incorporate and emphasize those aspects of the world that are important to the forecast. A review of scenario planning been presented in (Amur et al., 2013). Scenario-based design and evaluation for military capabilities has been analysed in (Urwinet et al., 2010). Scenarios are helpful in visualizing and understanding the incorporation of new systems within system of systems. The approach is based on the development of measure of effectiveness and performance and the techniques have been illustrated using cases that are relevant to network enabled capability. The measures of performance are independent of an operational scenario and allow the results to be compared with systems that provide the same functionality. In contrast, measures of effectiveness are dependent on an operational scenario. The scenario is composed of vignettes that contextualize the principal phases of systems development to meet a capability need. One or more operational vignettes must be included to test the deployed system (Urwinet et al., 2010). In our approach, three different representative scenarios have been used in the first questionnaire. The functionalities realised by system services and used in the operational tasks are similar in different scenarios. As a result, our second and third questionnaires which are not dealing with capability areas or system capabilities did not use scenarios explicitly. This is in accordance with (Urwinet et. al., 2010).

The practice of scenario planning implicitly accepts that managers are not able to make valid assessments of the likelihood of unique future events and that best guesses of what the future may hold may be wrong (Goodwin & Wright, 2010). Scenarios focus on key uncertainties and certainties about the future and construct vivid descriptions of the world. In their paper Goodwin and Wright review methods that aim to aid the anticipation of rare, high-impact, events. Methods are evaluated according to their ability to yield well-calibrated probabilities or point forecasts for such events. Authors conclude that all the methods are problematic for aiding the

anticipation of rare events and provide some remedies. Human judgement is often used to estimate the probabilities of events occurring. Goodwin and Wright point out possible cognitive biases. Events which are vivid, recent, unusual or highlighted by the media are assigned high probabilities. A tendency to ignore base-rate information and frequencies or anchoring on the current value and insufficient adjustment for the effect of future conditions are sources of biases in questionnaires based of human judgment. Our work is also based on questionnaires and human judgment, which should be paid attention when the results are assessed. However, our main objective has been to demonstrate the model building and for this purpose the questionnaire data is sufficient. Rare events occur in very special scenarios which make scenarios an essential tool for investigating rare events. Our method can be used directly for rare events and the results for capability areas and system capabilities can be used in planning and preparing for these rare events. Another way of using scenarios with different probabilities is to calculate an expected value for capability areas or system capabilities. The expected value is obtained with the probabilities of scenarios as weights in the sum of scenario capabilities.

A quantitative comparison between U.S. and U.S.S.R/Russian jet fighters (Bongers & Torres, 2013) estimates the relationship between the first flight date and a set of performance and technical characteristics such as thrust, climb rate, basic avionics, advanced avionics and stealth. Linear regression has been used as a mathematical tool. Another article includes a case study of main battle tank capability (Jiang et al., 2011). Weapon system capability assessment is a multiple criteria decision making problem with both quantitative and qualitative information under uncertain environment. Authors use belief structures model and evidential reasoning approach which were developed to deal with various types of uncertainties such as ignorance and subjectivity. The assessment framework for capabilities is hierarchic. A decision-analytic approach to reliability-based design optimization has been presented in a theoretical article (Bordley & Pollock, 2009). In their work, similar concepts to this paper are used but the approach is more theoretical. Several articles consider uncertain environments and use different methods for this purpose. Using probabilities might be suitable in many of these cases. Linear regression is not a optimal method for forecasting extreme events, such as the first flight day. As we have discussed in (Kuikka et al., 2015), few dedicated methods exist in this area. Because our model deals with interdependencies between technology areas, the method in this paper may be suitable also for these cases. Our method is a candidate for interpreting the first flight days, or comparable events. This can be considered in following studies.

A standardization of terminology of technology by conceptualizing products as complex artefacts that evolve in the form of a nested hierarchy of technology cycles. Such a system perspective provides both unambiguous definitions of dominant designs (stable core components that can be stable interfaces) and

inclusion of multiple levels of analysis (system, subsystems, components) (Murmann & Frenken, 2006). A new conceptual model for understanding technological evolution that highlights dynamic and highly interdependent relationships among multiple technologies has been proposed in (Adomavicius et al., 2005). The authors conclude that when technology evolution is discussed, a single technology cannot be considered in isolation. The technology ecosystem consists of a dynamic system that includes the totality of interrelated technologies. The authors identify three roles that technologies play within a technology ecosystem: components, products and applications and infrastructure. Types of interactions between technology roles have been classified as different kind of paths of influence. The use of the model has been demonstrated through the Wi-Fi business case. In this paper, because of the diversity of interdependencies between technologies, a simple matrix formulation is proposed for the modelling of technology areas. Modelling technologies with the principles of system of systems' methods, for example, is challenging because a technology area interacts with all the other technology areas.

A method for hierarchically prioritizing capabilities with an application to military manned and unmanned aerial vehicles provides the linkage between mission requirements to capability delivery options (Bourdon et al., 2014). The method provides a structure for breaking down requirements hierarchically into form against which selected capabilities can be assessed. Planning for military requirements is not bounded by a single threat that was expected to be faced in future conflicts. The planning is based on capabilities needed to win a conflict by defeating a range of threats encountered. The hierarchical prioritization of capabilities provides a broad high-level assessment of potential offered by different unmanned aerial vehicles. The method avoids comparing all the possible pairs of roles and tasks of UAVs as is necessary in the analytical hierarchy process (AHP). Another benefit of the model is that the addition of new missions, tasks, or roles does not invalidate previous assessments. This is not the case with other methods such as AHP in which the relative weights of two alternatives may change based on the introduction of a third alternative. Weapon selection using AHP and TOPSIS (technique for order performance by similarity to idea solution) methods under fuzzy environment has been analysed in (Dagdeviren et al., 2009). The AHP is used to analyse the structure of the weapon system in a fuzzy environment and to determine the wrights of the criteria, and fuzzy TOPSIS method is used to obtain final ranking. Again, the research problems in (Bourdon et al., 2014) and (Dagdeviren et al., 2009) bear resemblance with our research problems. Our method has also a favourable feature that adding a system to a system of systems does not result in recalculating all the pairs of systems.

Interoperability in military systems-of-systems architectures and capability-based quantitative technology evaluation have been addressed in (Wyatt et al., 2012) and (Biltgen & Mavris, 2007). The increasing complexity of net-centric warfare

requires systems to be interoperable to achieve mission success. The research surveys existing interoperability measurement methods and assess them from perspective of using interoperability as a metric to evaluate system-of-systems architectures. The purpose of the methodology has been to enable quantitative evaluation of technologies in a systems-of-systems context and enumerate how resources should be allocated to new technology development programs. The methodology provides insight into sensitivities of technologies on top-level capability metrics. In many cases, these sensitivities have been obscured by complexity of the problem. A holistic approach to the modelling and simulation of complex systems facilitates a traceable analysis process. The work demonstrates several ways that surrogate models can leverage to speed up processes, simultaneously examine technologies and tactics, and to enable next-generation visualization capabilities for systems-of-systems. The authors plan to extend their approach to examine multiple capabilities across a range of scenarios and will incorporate variable fidelity models to examine different trends and behaviours at varying levels of detail. Methods used in (Wyatt et al., 2012) and (Biltgen & Mavris, 2007) are different while the research goals and use of the results are similar. Simultaneous examination of technologies and tactics, resource allocation to new technology programs, sensitivity analysis of technology areas on top-level capability areas, system of systems' modelling and variable fidelity of systems and subsystems are common principles with our research methods.

Modelling the Interdependencies between Technology Areas

In this section we go through the steps of our model from technology areas to capability areas. Numerical results will be presented later in this paper, because we need some background from our earlier studies which will be presented in the next section. Our study is based on a questionnaire conducted in Finnish National Defence University in International What If?-data farming workshop. The questionnaire had two parts. In the first part the values for protection, situational awareness and engagement capability areas were asked for two systems and their combined use. The systems used were satellites and UAVs (Unmanned Aerial Vehicles).

In Figure 1 the structure of the overall model is described. In the first part (Q1 in Figure 1) of the questionnaire capability values for three capability areas were asked for satellites, UAVs and combined use of the systems. In the second part (Q2 in Figure 1) the effects of seven technology areas on five operational task capabilities were asked. Relationship "T/C" in Figure 1, is established by identifying three tasks with three capability areas. An additional survey (Q3 in Figure 1) was conducted among five technical persons (5 out of 10) of the same group. The analysis of Q3 results is the subject of this study. In this part,

interdependencies between technology areas were asked. The questionnaire structure is summarized in Table 1.

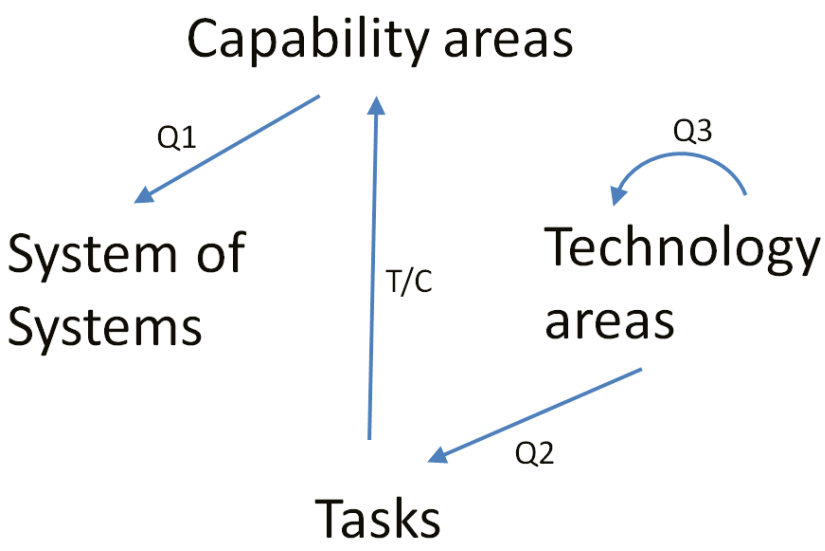

Figure 1. Hierarchic levels of the model. Three parts of the questionnaire are denoted Q1, Q2 and Q3. It is noticeable that there exists a sub model for each Qi (i=1, 2, 3) and the sub-models are modular components in the process flowchart.

In this paper our focus is in technological level. We present the connection between the development of technology areas and task capabilities. Our model consists of three phases corresponding three parts of the questionnaire. In our earlier work, we have presented models corresponding Q1 (Kuikka & Suojanen, 2014) and Q2 (Kuikka et al., 2015). In this paper we examine interdependencies Q3 between different technology areas and present the total process (Fig. 1) as well. With this knowledge, we are able to calculate capability changes caused by different technological changes. Results from part Q2 provide evaluated forecasting for 1 and 10 years. Combined with Q3, forecasting for less or more development of seven technology areas or combinations of different technology area developments can be calculated.

Technology areas used in the questionnaire are: sensor, material, communication, stealth, energy source, manufacturing and autonomous technology areas. We asked the technical persons consisting of five members of the original group of ten respondents, how great are the dependencies between different technology areas. The results are summarized in matrixes D_{UAV} and D_{SAT} in Appendix 1 for UAVs and satellites correspondingly.

Table 1. Three parts of the questionnaire Q1, Q2 and Q3.

	Description	Respondents	Time (T)
Q1	Changes in capability areas when satellites, UAVs or both have been deployed.	5 officers and 5 researchers	1 y, 10 y
Q2	Dependence of task capabilities on development of technology areas for satellites and UAVs.	5 officers and 5 researchers	10 y, 20 y
Q3	Interdependencies between technology areas.	5 researchers (same personnel as in Q1 and Q2)	10 y

In the first part of the questionnaire we have studied the relationship between capability areas and systems. In our previous work we have presented a model giving functional form between system capability values and capability area changes. For a capability area the system capability values can be expressed as functions of changes in capability areas values:

$$X_{SAT} = f_{SAT}(p_1, p_2, p_{12}) \text{ and}$$
$$X_{UAV} = f_{UAV}(p_1, p_2, p_{12}), \quad (1)$$

where X_{SAT} and X_{UAV} are system capability values with satellites and UAVs in use and p_1, p_2, p_{12} are changes of capability areas values with satellites, UAVs and combined use of satellites and UAVs. Changes are measured between the initial values p_0 (T=0) of capabilities and the final values (T=1, 10).

We present two possible alternatives for the functional forms in Equation (1). Our own model from (Kuikka & Suojanen, 2014) is shortly summarized with the following equations:

$$\mathrm{p}_0 = \mathrm{X}_\mathrm{m}\mathrm{X}_\mathrm{k}$$
$$\mathrm{p}_0 + p_1 = \mathrm{X}_\mathrm{m}(1 - (1 - \mathrm{X}_\mathrm{SAT})(1 - X_k))$$
$$\mathrm{p}_0 + p_2 = \mathrm{X}_\mathrm{m}(1 - (1 - \mathrm{X}_\mathrm{UAV})(1 - X_k))$$
$$\mathrm{p}_0 + p_{12} = \mathrm{X}_\mathrm{m}(1 - (1 - \mathrm{X}_\mathrm{SAT})(1 - X_{UAV})(1 - X_k)), \quad (2)$$

where X_{SAT}, X_{UAV} are the system capability values for satellites and UAVs. And X_m and X_k are the system capability values for serial and parallel systems functioning with the satellite and UAV systems. Equations (2) follow from the probabilistic definition of capabilities. In our basic model the equation are similar for the three capability areas under consideration: protection, situational awareness and engagement.

From Equations (2) we can solve X_{SAT}, X_{UAV}, X_k, and X_m:

$$X_{SAT} = \frac{p_1 + p_2 - p_{12}}{p_2}$$

$$X_{UAV} = \frac{p_1 + p_2 - p_{12}}{p_1}$$

$$X_m = p_0 + \frac{p_1 p_2}{p_1 + p_2 - p_{12}}$$

$$X_k = (1 + \frac{p_1 p_2}{p_0 (p_1 + p_2 - p_{12})})^{-1} \tag{3}$$

Serial and parallel systems need not be specified in more detail, however their capability values X_m and X_k can be used in the analysis if desired. More details of the model (Q1) are presented in (Kuikka & Suojanen, 2014).

Martino (Martino, 1993 and Kim, 2012) introduced a scoring model for rating technology quantitatively:

$$S = \frac{A^a B^b (cC + dD + eE)^z (fF + gG)^y (1 + hH)^x}{(iI + jJ)^w (1 + kK)^v}, \tag{4}$$

where $c+d+e=1$, $f+g=1$, $i+j=1$, $a+b+z+y+x=1$ and $w+v=1$. In the model A and B are overriding factors and {C, D, E}, {F, G}, and {I, J} are exchangeable factors within brackets. I, J, and K are costs or undesirable factors. The factors $(1 + hH)$ and $(1 + kK)$ represent special cases that must stand alone but cannot be traded off with any other factors. Moreover, they may not always be present. Factor H is not overriding, in the sense that its absence justifies a score of 0. It is an option that increases the score if present but does not affect the score if absent. In the same way, just because undesirable factor K has a value of 0 for some devices does not mean that their score should be infinite. In Equation (4) h and k are constants. The use of this method is an exception to the rule that weights must be normalized to sum to 1.0. Since there is only one factor in each group, however, normalization does not distort the overall score (Martino, 1993). Martino's model can be compared with our model, the serial system X_m can be considered a combination of overriding factors and the parallel systems are exchangeable factors in Martino's formulation. Cost or undesirable factors are not examined explicitly in our model, their effects are included in evaluated capability values. In our method, systems of systems are modelled with serial and parallel systems and no extra factors like $(1+hH)$ exist. Another scoring model with similar ideas has been introduced by Gordon and Munson (Martino, 1993 and Kim, 2012).

From Figure 1, we see that following the chain "Q3-Q2-T/C-Q1" changes in capability areas and systems capabilities can be calculated as a function of technology area developments. The overall model is modular in the sense that in the modelling process sub models Q1, Q2, Q3 or T/C can be changed. For example, in this paper the model for T/C dependency is the simplest choice of just identifying operational tasks with capability areas. A refined model could be a weighted combination or a function of different task capabilities. Another example would be a more detailed system model in Q1 resulting in different functional forms of f_{SAT} and f_{UAV} in Equations (1) and (3). In other words, we have constructed the method in the spirit of system of systems. The idea of the chosen approach is to distinguish the task into two parts: firstly to form an information model for the existence of dependencies; and secondly to construct the detailed models for the dependency structures. The purpose of this paper is to draw together the individual models.

Next we present the relationship between the development of technology areas and task capabilities. In our model, task capabilities are calculated as average values of technological development values:

$$C_j = \frac{1}{7}\sum_{i=1,7} A_{i,j},\ j = 1,...,5, \tag{5}$$

where matrix element $A_{i,j}$ describes the development of technology area i on task capability j. Justification for the average value as a measure is that we regard the respondents automatically weighted the importance of each technology area affecting task capabilities. Sum of technology development values is normalized, and as a consequence, we end up with the average value in Equation (5).

We need matrices from second (matrix A) and third part (matrix D) of the questionnaire. Matrix A gives the relationship between technology areas and tasks. Matrix D gives the interdependency between technology areas. In Equation (6) $\Delta_k,\ k = 1,...,7$ describes development in technology area k less or more than forecasted in the questionnaire data. The forecasted values are listed later in Table 5. For example, if $\Delta_1 = 0$ no development in sensor technology occurs, and if $\Delta_2 = 1.5$ material technology develops 50 % better than forecasted. To make this clear, parameter values of Δ_k are listed in Table 8 in Appendix 1.

$$C_j = \frac{1}{7}\sum_{i=1,7} A_{i,j}\left(1 + D_{i,k}\left(\Delta_k - 1\right)\right),\ j = 1,...,5,\ k = 1,...,7. \tag{6}$$

In Equation (6) we assumed that only one of the parameter values Δ_k, at a time, is different from one: $\Delta_k \neq 1$. We will relax this assumption later in Equation (8). In Equation (6), if $\Delta_k = 1$, the development of technology area *k* is at the same level as forecasted in Table 5. Matrix element $D_{i,k}$ describes the effect of technology area *k* on technology area *i*. For simplicity, we omit index *k* on the left side of the equation. Next, we calculate a special case of task capabilities where one of the technology areas has no development. In this case $\Delta_k = 1, k = 1,...,k' - 1, k' + 1...,7$ and $\Delta_{k'} = 0$ (for $k = k'$). In Equation (6) the term inside the sum is:

$$\begin{aligned} &1 + D_{i,k}(\Delta_k - 1) = 1 - D_{i,k'} \quad \text{for } j = 1,...,5 \text{ and } k = k' \\ &1 + D_{i,k}(\Delta_k - 1) = 1 \quad \text{for } j = 1,...,5 \text{ and } k \neq k'. \end{aligned} \tag{7}$$

Next, Equation (6) is further generalized in Equation (8). We can calculate iteratively task capabilities when more than one technology area develops less or more than evaluated in Table 5.

$$C_j = \frac{1}{7} \sum_{i=1,7} A_{i,j} \prod_{k=1,7} \left(1 + D_{i,k}(\Delta_k - 1)\right), \; j = 1,...,5, \tag{8}$$

where the product is taken over all technology areas. In the previous section, we discussed possible extensions of the model in Equation (8). Jumps, triggered at specified levels of capabilities or at some other conditions, can easily be implemented, for example, in a spreadsheet application. Also, we discussed the applications for rare events and expected value evaluations.

In Equations (6) and (8) we assume that technology development is linear having no "quantum leaps" when $\Delta_k \neq 0,1, k = 1,...,7$. These disruptive technological changes may cause dramatic impacts on other technology areas and enable or disable certain capabilities. Also the linear improvement (e.g. capacity of energy storage) does not necessarily correspond linear improvement on the capabilities, but there may be thresholds before any improvements appear. Besides, negative impacts of achievements of the technology areas on the capabilities are possible, for example, improvements in stealth technology or materials may decrease capability to perform surveillance as effectively as today. However, we reported in our first conference article (Suojanen et al., 2014): "Technology development of the RED nation was frozen in the estimation of the capabilities, since the focus was on own assets." Consequently, no negative impacts caused by the adversary's technological counter measures exist in the results.

Research Data and Error Analysis

In this section, we present research data from questionnaires Q1 and Q2 in the extent that is needed for analysing the results of questionnaire Q3. Presentation of new results of this paper is postponed to the next section. At the end of this section an error analysis is conducted. The relationship between the score values and the percentage values from questionnaires Q1 and Q2 is assumed to be valid also in questionnaire Q3. This can be justified because five of the experts participated in both questionnaires Q2 and Q3 (see Table 1). Because the data from our earlier work is used, we present the error analysis in this section. The same error levels are assumed to hold in questionnaire Q3.

Two questionnaires (Q1 and Q2 in Figure 1) have been conducted in our earlier research (Kuikka & Suojanen, 2014 and Kuikka et al., 2015). In Table 2 the average capability changes for situational awareness from the questionnaire Q1 are shown. The average value is justified because no scenarios were explicitly used in Questionnaire 2 and at the same time the evaluations were giving on the grounds of three basic scenarios from Questionnaire 1. Another justification is provided in (Urwin et al., 2010) where scenarios are used only in effectiveness measures, not in performance measures. Note that the changes are provided at the level of capability areas, not for satellite or UAV systems. The initial values for the awareness capabilities are 0.8, 0.6 and 0.4 for Scenarios 1, 2 and 3 respectively. For example, in Scenario 1 after ten years development the awareness capability is 0.8 + 0.085 = 0.885 (88.5 %) when UAVs are in use. In Reference (Kuikka & Suojanen, 2014) also results for protection and engagement capability areas are given for 1, 10 and 20 years' time span. In this paper we study situational awareness 10 years development as an example, other cases can be examined similarly. The initial values for protection, awareness and engagement capability areas together with average capability changes for 10 years development are summarized in Table 3.

Table 2. The changes of situational awareness capability produced by satellite and UAV systems for Scenarios 1, 2 and 3 in 10 years. The last column shows the average values of scenarios.

Situational Awareness	10 years			
	Sce1	Sce2	Sce3	average
SAT	0.063	0.083	0.157	0.101
UAV	0.085	0.123	0.173	0.127

Table 3. The initial capability values (T=0) for Scenarios 1 – 3 are shown. The average capability value changes with satellites and UAVs are calculated for 10 years development. The average capability changes have been calculated similarly as the average values for situational awareness in Table 2.

Capability area	Scenario 1 T=0	Scenario 2 T=0	Scenario 3 T=0	Average SAT, T=10	Average UAV, T=10
Protection	0.7	0.5	0.5	0.071	0.087
Sit. Awareness	0.8	0.6	0.4	0.101	0.127
Engagement	0.9	0.3	0.4	0.049	0.101

As can be seen from Table 1, both Questionnaire 1 and Questionnaire 2 include evaluations for 10 years technological development. In addition, the evaluated changes have been given in two different formats, in percentages and in score values 0 – 3. The same personnel took part in both questionnaires. These facts give us an opportunity to formulate a functional relationship between scores and percentages. The functional form of the dependence between the score values and the percentage values must be linear, since they describe the same quantity, only the scaling is different. Figure 2 shows the correspondence together with the linear regression fit 7.58*Score-1.39 between the two data sets. In spite of the fact, that Questionnaire 1 and 2 have been conducted one after the other, because of different formats and limited time for answering, the questionnaires can be considered almost independent of each other. As a result, we get a shortcut for estimating error levels of capability changes. The 95 % confidence interval (Levine et al., 2010, Chapter 13, Equation 13.20) for the forecasted mean value 8.9 is (8.9-1.5, 8.9+1.5)=(7.4, 10.4). Another statistic is the coefficient of determination $R^2 = 0.80$ and the corresponding correlation coefficient R = +0.89. These statistics indicate that the linear relationship gives a fairly good description of the data. The matrix element values A_{UAV} and A_{SAT} in Appendix 1 and the corresponding values in Table 5 have been calculated from the linear relationship.

Figure 2 gives us a practical perception about error levels. Because of subjective evaluations and the small number of participants of the questionnaires no detailed analysis, and no detailed error analysis, is meaningful. After all, our main purpose is to demonstrate methods of model building. In Reference (Kuikka et al., 2015) we have presented standard deviation values of the questionnaire data. The results are consistent with the confidence level and the coefficient of determination statistics. The average capability values and the corresponding standard deviations of ten respondents are summarized in Table 4. The standard deviations are considerably higher compared to the previous confidence level estimation. The explanation is that they measure different uncertainties. Standard deviations describe individual points and higher standard deviations are a consequence of different levels of

estimation among respondents. In the regression analysis, the confidence level describes the goodness of the regression line to predict the linear relationship between the two data sets. The confidence level approach is more appropriate for the situation at hand.

Table 4. The average Score values from Questionnaire 2 and the corresponding standard deviations (STD) in parenthesis.

	SAT, T=10 Score, (STD)	UAV, T=10 Score, (STD)
Deception (Prot)	1.11 (0.87)	1.53 (0.92)
Surveillance (Awa)	1.27 (0.88)	1.76 (0.74)
Engagement (Eng)	0.91 (0.83)	1.59 (0.86)

Figure 2 shows that the confidence interval close to the mean point (1.5, 10.0 %) is about 2.0 % and 4.0 % for the 90 % and 97.5 % confidence levels respectively. The confidence interval is wider for the values not close to the mean point. The confidence intervals can be used to observe the significant values in Tables 6 and 7. The last two columns in the tables show the 90 % and 97.5 % confidence level values (½*interval) and the significant changes on the confidence level 90 % are bolded in the Tables 6 and 7. The results indicate that the UAV capability changes are statistically significant on the 90 % confidence level except for the engagement capability change. The satellite capability changes are significant for the surveillance and communication capability changes on the 90 % confidence level except for the stealth technology. For the deception capability change the communication and autonomous technologies are statistically significant. These results are intuitively very understandable. (For comparison, the last column in Tables 6 and 7 show the 97.5 % confidence level thresholds. Interestingly, on the 97.5 % confidence level the surveillance capability changes for UAVs are not statistically significant.)

The uncertainties in matrix elements $D_{i,k}$ should be analysed with the help of standard deviations. Because of a small number of respondents (5) the standard deviations are very high. The standard deviations are $2^{1/2}$ times higher compared to the values in parentheses in Table 4. We can only make a rough estimate about the uncertainties: the errors are 30 % on the basis of one unit standard error (error in Tables 6 and 7 values is roughly 0.005).

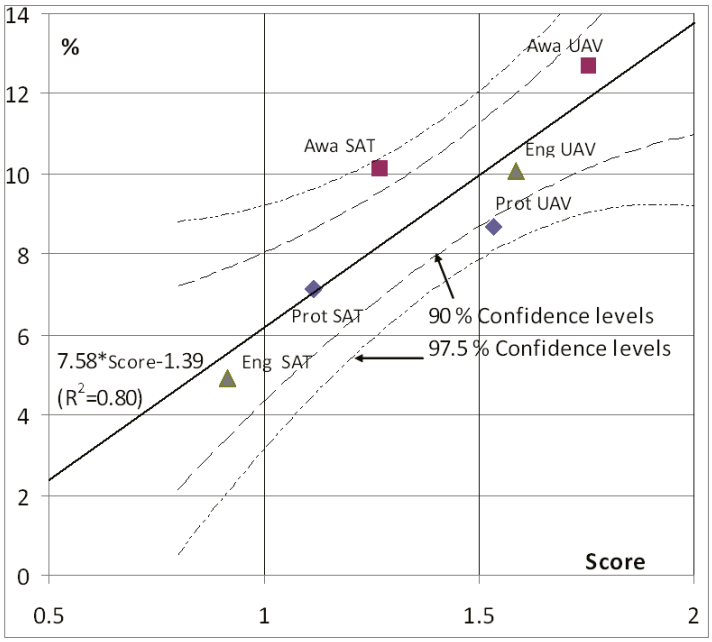

Figure 2. The relationship between Questionaire 1 (%) and Questionaire 2 (Score) data. The regression line and the 90 % and 97.5 % confidence levels are shown. The full set of parameters of the statistical analysis is presented in Appendix 2.

Numerical Results of This Study

The theoretical background has been presented earlier in this paper in section "Modelling the Interdependencies between Technology Areas". In this section the numerical results and some examples of applying the theory for different technology areas are presented. We show also how the results of this study can be combined with the results of our earlier studies.

In Table 5, the capability changes for UAVs and satellites for five operational tasks are calculated from Equation (5). The values correspond to the linear regression fit of Figure 2.

Table 5. The capability changes for UAVs and satellites from Equation (5).

	Surveillance	Communications	Engagement	Logistics	Deception
UAVs	0.119	0.107	0.106	0.082	0.102
Satellites	0.082	0.094	0.055	0.051	0.071

In Tables 6 and 7, changes downward in percentages are given for UAVs and satellites (Note that because the model for interdependencies is linear, changes upwards are considered similarly). We illustrate the use of the tables by an example. According Table 6, if UAV sensor technology is not developing at all,

while other UAV technologies develop as evaluated in Table 5, surveillance task capability of UAV system decreases from 11.9 % (Table 5 UAV Surveillance) to 9.1 % (11.9 % - 2.8 % = 9.1 %). This stands for 23.2 % relative decrease. The situational awareness value is the sum of the initial capability value (Table 3) and the change, for example in Scenario 1 the capability value is 85.2 % (0.8+0.085*(11.9/12.7)-0.028=0.852). Here, we have taken into account the adjusting between the regression line estimate 11.9 % (in Table 5) and the questionnaire result 12.7 % (in Tables 2 and 3), as we have concluded that the regression line approach eliminates some sources of error in the evaluations.

We provide one example how the errors can be evaluated. Based on the estimations in the previous section, the error estimation is roughly 0+0.2*0.085+0.3*0.028=0.025. The initial value has no error because the value 0.8 is defined in the scenario. As a result, the situational capability area value for Scenario 1 UAV systems in use is between 83 % - 85 % (82.7 % - 85.2 %) with no development in sensor technology. Other cases can be evaluated similarly.

Table 6. The decrease of task capabilities when no development occurs on one of the technology areas (UAVs).

Task/ Technology	sensor	material	comm	stealth	energy	manuf	autom		$CL_{90\%}$	$CL_{97.5\%}$
Surveillance	**0.028**	**0.029**	**0.026**	**0.025**	**0.027**	**0.027**	**0.025**		0.019	0.032
Comm	**0.024**	**0.026**	**0.027**	**0.020**	**0.025**	**0.024**	**0.022**		0.015	0.024
Engagement	0.023	0.027	0.022	0.025	0.021	0.027	0.022		0.014	0.024
Logistics	**0.015**	**0.022**	**0.018**	**0.013**	**0.020**	**0.023**	**0.018**		0.012	0.020
Deception	**0.023**	**0.028**	**0.020**	**0.023**	**0.017**	**0.025**	**0.026**		0.013	0.022

Table 7. The decrease of task capabilities when no development occurs on one of the technology areas (satellites).

Task/ Technology	sensor	material	comm	stealth	energy	manuf	autom		$CL_{90\%}$	$CL_{97.5\%}$
Surveillance	**0.022**	**0.017**	**0.019**	0.007	**0.019**	**0.020**	**0.016**		0.012	0.020
Comm	**0.021**	**0.022**	**0.023**	0.010	**0.022**	**0.021**	**0.017**		0.012	0.019
Engagement	0.014	0.013	0.012	0.006	0.014	0.015	0.012		0.021	0.035
Logistics	0.009	0.015	0.009	0.007	0.011	0.014	0.009		0.023	0.038
Deception	0.014	0.013	**0.019**	0.010	0.013	0.014	**0.018**		0.015	0.025

Analysing the values in Tables 6 and 7 gives an understanding of the level of dependency between technology areas and task capabilities when interdependencies between different technology areas have been taken into account. For example, UAV material and sensor technology has great influence on

surveillance. This is an understandable result when we consider how the surveillance task is already enhanced by e.g. SAR (Synthetic Aperture Radar).

For UAV we observe also how the lack of development in material technology has a great effect on all the tasks except logistics. For satellites one observes how the logistics task is not sensitive on the development of sensor technology. It is interesting to see how the lack of development on stealth technology does not affect greatly to any of the UAV tasks and affects weakly to the satellite tasks. In general, the lack of development in any technology area affects more on the development of UAV than satellite capabilities.

If we examine the rows of Tables 6 and 7, we observe two main results. One is that UAV surveillance task is dependent on the development of the majority of the considered technology areas. The other is that the logistics task of UAV is weakly dependent on the majority of the technology areas development and almost independent for satellites.

In the following, we examine surveillance task. Similar results can be obtained for communication, engagement, logistics and deception operational tasks. Figures 3 and 4 show examples of calculations where multiple technology areas have different development behavior than evaluated in Table 5. In Figure 3, less than evaluated technological development, while in Figure 4 more than evaluated technological development, is occurring. Figure 3 is for UAV technologies and Figure 4 shows both UAV and satellite technologies for comparison. Curves are cumulative from left to right: curves at point “energy” have less development in energy technology and evaluated development (Table 5 values) in other technology areas, curves at point “sensor” have less development in “energy” and “sensor” and evaluated development in “comm”, “auto”, “stealth” and so on.

The choices 50 % and 10 % are representative examples for great and small deviations from the evaluated development. Any combinations of $\Delta_k, k = 1,...,7$ and their values can be calculated from Equation (8). The order of cumulative calculation is a choice representing a reliability of a system’s functioning which is another concept than dependency of technological developments. Energy, sensor and communication technologies are the most critical in this respect while material and manufacturing technologies have more profound influence on capabilities because of their enabling role with other technologies. Again, any order, combination or different values can be calculated from Equation (8). In calculating the simultaneous effects of multiple technologies the order of technologies does not change the results in the linear model of Equation (8). Service and technical characteristics is another classification of technologies (Windrum et al., 2009) and from the user point of view services appear more important than purely technical characteristics.

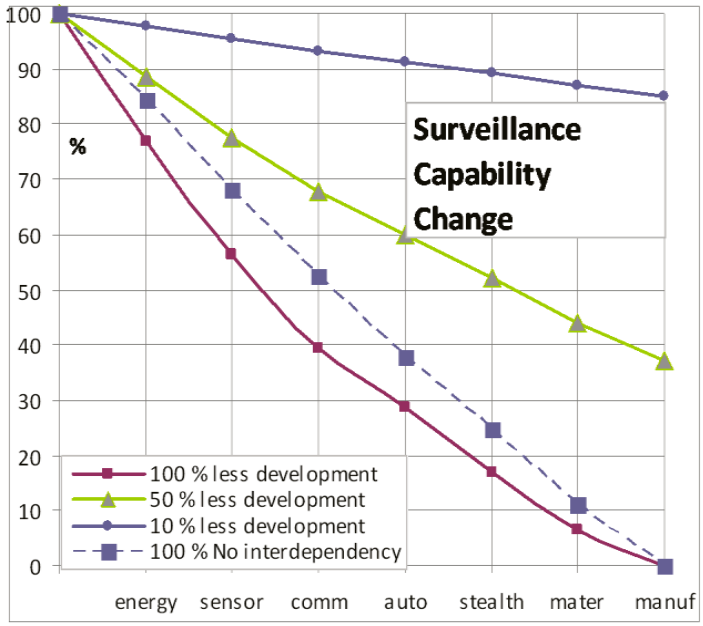

Figure 3. An example showing decreasing surveillance capability when more than one technology area has no development (100 % decrease), 50 % less or 10 % less development than evaluated in the questionnaire. Curves are for UAVs. As a comparison, the dotted curve for 100 % decrease has no interdependency between technology areas.

In Figure 3, the dotted curve shows the case where no interdependencies exist between technology areas. The effect of interdependencies can be more than 10 % when 100 % less (no development) development in more than one technology areas occurs. At point "autonomous technology" none of the technology areas have additional or less development compared to Table 5, so interdependency has no effect. If all technology areas have technology development as evaluated in Table 5, interdependency is already implicitly included in the data.

The application of Equation (8) and Figures 3 and 4 is when unexpected technical inventions and developments change existing information about developments in technology areas. On the other hand Equation (8) can be used as a method for sensitivity analysis of one or several technology areas. The curves 100 % in Figure 4 are only for a sensitivity analysis, because the linear equations are not valid for very large values. From the changed technology developments the effects on capabilities can be calculated as presented in (Kuikka & Suojanen, 2014, Suojanen et al., 2015 and Kuikka et al., 2015).

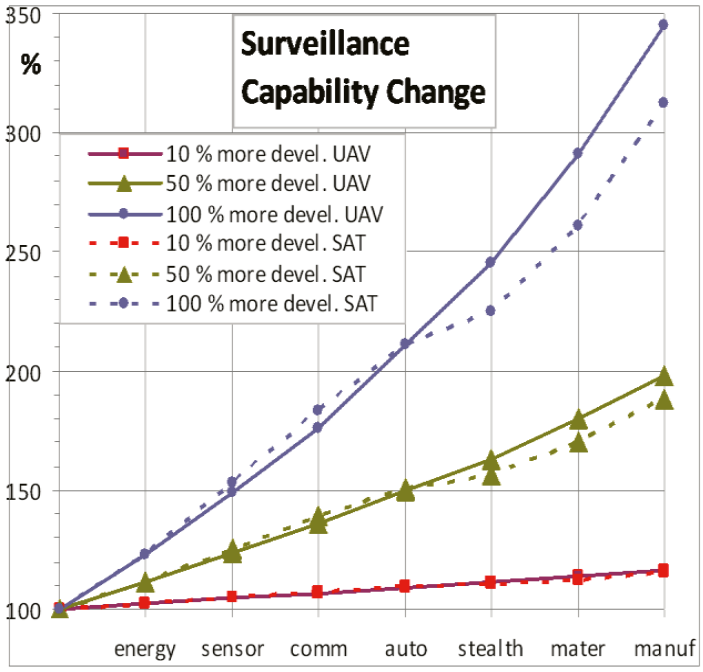

Figure 4. Same as Figure 3 but more development than evaluated in the questionnaire has been assumed. Curves are cumulative as in Figure 3. Solid curves are for UAVs and dotted curve are for satellites.

In Figure 5, we demonstrate the use of Equation (8) in a more complicated situation. In practice, this means we can allow any values for Δ_k we want and brings us closer to determining the total uncertainty (notice analogy to measurement uncertainty analysis). Developments in technology areas deviate from the initial values in Table 5 according $\Delta_k, k = 1,...,7$. For example, the first curve has $\Delta_1 = 0.8, \Delta_2 = 1.5, \Delta_{3,4,5,6,7} = 1.1$. Again, the curve shows cumulative effect of $\Delta_k, k = 1,...,7$ from left to right. Solid curves are for UAVs and dotted curves are for satellites. In Figure 5, significant differences in UAV and satellite surveillance capabilities can be observed. For example, satellites are more sensitive to sensor technologies and UAVs are more sensitive to stealth technologies.

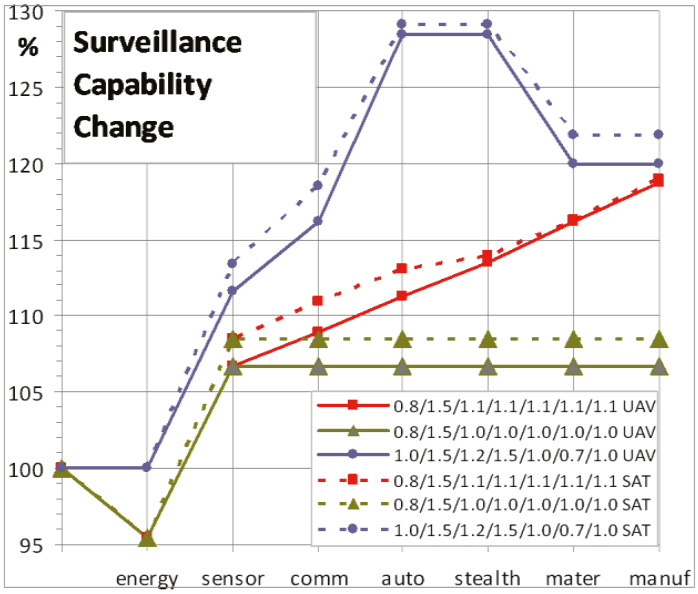

Figure 5. Examples of using the model with different technology area developments. In figure values of Δ_j, $j = 1,...,7$ in Equation (8) are shown. Curves are cumulative in Δ_j. Solid curves are for UAVs and dotted curves for satellites. Values of Δ_j, $j = 1,...,7$ are shown for each curve.

Conclusions

In this article, we have presented a process flowchart demonstrating how the developments in technology areas end up to capability area development (and beyond). We formulated mathematical methods in analyzing the intermediate steps of the process. While all the numerical results are computed by the methods in this / these article(s) it is important to notice that one is able to use alternative methods also; one of our main results in this work is to demonstrate the process itself.

On the system level our model makes use of system of systems principles. Satellites and UAVs have been assembled in parallel or in series with other systems. This idea can be compared with the ideas in Martino's more heuristic model. Other models, such as the analytical hierarchy process (AHP), require all the weights to be recalculated after a new system or task is added into the calculation. In our method, if no significant secondary effects exit, a new system can be added in serial or in parallel to an existing set of systems with a limited amount of recalculations (Kuikka & Suojanen, 2014).

We presented also a tool for estimating the interdependencies between technology areas. The numerical results are obtained from a questionnaire asked from 5 researchers. Results of the questionnaire have been presented for the first time in this paper. We show how the proposed method may be applied in conducting

sensitivity analysis of the technology development. Our vision is to raise the sensitivity analysis of the technology development to the level that it has e.g. in the measurement technology. In this paper, we provide several examples of the use of the model for technological interdependencies and sensitivity analysis.

One is able to use the proposed method in analyzing the effects of unexpected technical inventions and developments. The analysis can be conducted by knowing the interdependencies of technology areas and to adjust the knowledge with the updated situation (e.g. more effective sensors etc.). Only rough conclusions can be made from results of the three questionnaires because of a small number of participants and uncertainties in human judgment. Nevertheless, the results can be used when the limitations are taken into account. The approach can be used in illustrating general impacts of various technological areas and their interdependencies. This can be used, for example, in balancing resource allocation between technological areas. The results show that material technology has the most widespread influence on other six technology areas examined. UAV material and sensor technology has a great influence on surveillance. This is an understandable result when we consider how the surveillance task is improved by radars and other sensors. Logistics task is not sensitive on the development of sensor technology in satellites. It is interesting to see how the lack of development on stealth technology does not affect greatly to any of the UAV tasks and affects weakly to the satellite tasks. In general, the lack of development in any technology area affects more on the development of UAV than satellite capabilities. As said, questionnaire data has been examined with the model and methods of this paper. More realistic results may be obtained with system modelling or simulation results as input data instead of questionnaires.

We demonstrated how to extrapolate impacts of technology changes on capabilities with a linear model. This does not take into account possible disruptive or threshold effects. If a threshold and its impacts on the technological areas are known a non-linear variant of the model is a straightforward extension of our work. On the other hand, disruptive changes are unpredictable and their timing and effects are difficult or impossible to model.

In our literature review we have made some observations about similarities and differences of methods and models when comparing with our approach. Modelling capabilities is usually based on scoring or other measures with no quantitative interpretation. Our model has an exact interpretation: capability is defined as the probability of success. In the modelling we have concluded that a minimal taxonomy includes capability areas, systems of systems and technology areas. These all have their internal models and interdependencies. Especially, modelling technological interdependencies appears to be a complex task. A basic linear deterministic model has been presented in this paper. The model can be compared with the stochastic correlation matrix of statistics. The correlation matrix is

symmetric while the model proposed in this paper allows different values for dependencies depending on the direction of the association.

In our modelling scenarios are used in modelling capability areas while scenarios are not used in modelling technological areas. This simplifies the questionnaires and the modelling efforts. A comparable approach has been presented in (Urwin et al., 2010) where the measures of performance are independent of an operational scenario and allow the results to be compared with systems that provide the same functionality.

References

Gediminas Adomavicius, Jesse C. Bockstedt, Alok Gupta, Robert J. Kauffman (2005). Technology roles and paths of influence in an ecosystem model of technology evolution, *Information Technology and Management*, Vol. 8, Issue 2, 185-202.

Muhammad Amer, Tugrul U. Daim, Antonie Jetter (2013). A review of scenario planning, *Futures* 46, 23-40.

Patrick T. Biltgen, Dimitri N. Mavris (2007). *Capability-based quantitative technology evaluation for systems-of-systems*, IEEE System of Systems Engineering Conference.

Anelí Bongers, José L. Torres (2014). Measuring technological trends: A comparison between U.S. and U.S.S.R./Russian jet fighter aircraft, *Technological Forecasting & Social Change*, Vol. 87, 125-134.

Robert F. Bordley, Stephen M. Pollock (2009). A decision-analytic approach to reliability-based design optimization, *Operations Research* 57, 1262-1270.

Sean Bourdon, Benoit Arbour, Matthew R. MacLeod (2014). A method for hierarchically prioritizing capabilities with an application to military manned and unmanned aerial vehicles, *Journal of Applied Operational Research* 6(1), 39-47.

Carolina Castaldi, Roberto Fontana, Alessandro Nuvolari (2009). The evolution of tank technology, 1915-1945, *Journal of Evolutionary Economics* 19(4), 545-566.

Mario Coccia (2003). *An approach to the measurement of technological chance based on the intensity of innovation*, Ceris working paper.

Mario Coccia (2005). Technometrics: Origins, historical evolution and new directions, *Technological Forecasting & Social Change* 72, 944-979.

Metin Dagdeviren, Serkan Yavuz, Nevzat Kilinc (2009). Weapon selection using the AHP and TOPSIS methods under fuzzy environment, *Expert Systems and Applications* 36, 8143-8151.

Jan van den Ende, Wilfred Dolsma (2002). *Technology push, demand pull and the shaping of technological paradigms – patterns in the development of computing technology*, ERIM report series ERS-2002-93-ORG.

Koen Frenken, Loet Leydesdorff (2000). Scaling trajectories in civil aircraft (1913-1997), *Research Policy* 29(3), 331-348.

Koen Frenken (2006). Technological innovation and complexity theory, *Economics of Innovation and New Technology* 15(2), 137-155.

Paul Goodwin, George Wright (2010). The limits of forecasting methods in anticipating rare events, *Technological Forecasting & Social Change* 77, 355-368.

Jan G. De Gooijer and Rob J. Hyndman (2006). 25 Years of Timer Series Forecasting, *International Journal of Forecasting*, 22 (3), 443–473.

Theodore J. Gordon, Thomas R. Munson (1981). A proposed convention for measuring the state of the art of products or processes, *Technological Forecasting & Social Change* 20, 1-26.

Juhani S. Hämäläinen, Juha-Pekka Nikkarila (2015). *Analyzing Strategy Effectiveness With Resource Profit Ratio Integrals*, ISMS Conference.

Andrew D. James (2013). *Emerging technologies and military capability*, RSIS / Policy Briefs, Retrieved 10.10.2015 from http://www.rsis.edu.sg/search/?q=andrew+james.

Jiang Jiang, Xuan Li, Zhi-jie Zhou, Dong-ling Xu, Ying-wu Chen (2011). Weapon system capability assessment under uncertainty based on the evidential reasoning approach, *Expert Systems with Applications* 38, 13773-13784.

Byoung Soo Kim (2012). Measuring technological change – concepts, methods, and implications in *Technological Change*, Aurora Teixeira (Ed.), Retrieved 10.10.2015 from http://www.intechopen.com/books/howtoreference/technological-change/measuring-technological-change-concept-methods-and-implications.

Vesa Kuikka, Juha-Pekka Nikkarila, Marko Suojanen (2015). *A technology forecasting method for capabilities of a system of systems*, PICMET Conference.

Vesa Kuikka, Marko Suojanen (2014). Modelling the impact of technologies and systems on military capabilities, *Journal of Battlefield Technology*, Vol. 17, No. 2, 9-16.

David M. Levine, Mark L. Berenson, Timothy C. Krehbiel, David F. Stephan (2010). *Statistics for Managers Using MS Excel*, 6th Edition, Pearson, Chapter 13 Retrieved 10.10.2015 from http://www.prenhall.com/behindthebook/0136149901/pdf/Levine_CH13.pdf.

Joseph P. Martino (1993). *Technological Forecasting for Decision Making*, McGraw-Hill, Inc., McGraw-Hill Engineering and Technology Management Series.

Johann Peter Murmann, Koen Frenken (2006). Toward a systematic framework for research on dominant designs, technological innovations, and industrial change, *Research Policy* 35, 925-952.

Pier Paolo Saviotti, Adreas Pyka (2013). The co-evolution of innovation, demand and growth, *Economics of Innovation and New Technology* 22(5), 461-482.

P. P. Saviotti, G. S. Mani (1995). Competition, variety and technological evolution: a replicator dynamics model, *Evolutionary Economics* 5, 369-392.

Antoine Schoen, Lionel Villard, Patricia Laurens, Jean-Philippe Cointet, Gaston Heimeriks, Floortje Alkemade (2012). *The network structure of technological development; technological distance as a walk on the technology map*, Science & Technology Indicators (STI).

Johan Schot, Frank W. Geels (2007). Niches in evolutionary theories of technical change – A critical survey of the literature*, J. Evol Econ*, Vol. 17, 605-622.

Seunghwan Oh, Sungmoon Jung, Jaemin Park, Chang Sun Kim (2013). Measuring technology level on a mixed model: A case of Korean photovoltaic technology, *International Journal of Digital Content Technology and its Applications*, Vol. 7, No. 13.

Jiyeon Ryu, Soon Cheon Byen (2011). Technology level evaluation methology based on the technology growth curve, *Technological Forecasting & Social Change* 78, 1049 – 1059.

Marko Suojanen, Vesa Kuikka, Juha-Pekka Nikkarila, Jari Nurmi (2015). *An example of scenario-based evaluation of military capability areas – an impact assessment of alternative systems on operations*, IEEE International Systems Conference.

Esmond N. Urwin, Colin C. Venters, Duncan J. Russell, Lu Liu, Zongyang Luo, David E. Webster, Michael Henshaw, Jie Xu (2010). *Scenario-based design and evaluation for capability*, 2010 5th International Conference on System of Systems Engineering.

Paul Windrum, Cecilia Diaz, Despoina Filiou (2009). Exploring the relationship between technical and service characteristics, *J Evol Econ* 19 567-588.

Elizabeth Jones Wyatt, Kelly Griendling, Dimitri N. Mavris (2012). *Addressing interoperability in military systems-of-systems architectures*, Systems Conference.

Charles Zaiontz (2015). Real Statistics Using Excel, Retrieved 10.10.2015 from http://www.real-statistics.com.

Appendix 1

We have used the same relationship between scores and probability values (capabilities) as in our previous paper:

$$Capability = 0.0758*Score-0.0139. \quad (9)$$

Scores 0, 1, 2 and 3 have been used in evaluation because probability values are more difficult to estimate. Equation (9) has been derived from the first and second parts of the questionnaire with the help of the fact that the same answers have been given as probabilities and as scores in these two parts of the questionnaire. In Table 1, these are the cases for satellites and UAVs for 10 years development with three capability areas identified with corresponding three tasks: protection with deception task, situational awareness with surveillance and engagement with engagement task. This is justified by the fact that in the second part of the questionnaire general tasks were evaluated with no direct connection with scenarios of the questionnaire. In other words, results concerning task capabilities are averages over three scenarios of the questionnaire.

Matrixes used in calculations for UAVs and satellites. Matrix A gives the relationship between technology areas and tasks. Matrix D gives the interdependency between technology areas. Equation (9) is used in transformation from scores to probabilities (capabilities). Note that the columns and rows are arranged in the same order than in Tables 6 and 7.

$$A_{UAV} = \begin{pmatrix} 0.138 & 0.115 & 0.107 & 0.062 & 0.107 \\ 0.115 & 0.107 & 0.107 & 0.100 & 0.123 \\ 0.130 & 0.145 & 0.107 & 0.085 & 0.087 \\ 0.107 & 0.077 & 0.123 & 0.039 & 0.107 \\ 0.130 & 0.123 & 0.092 & 0.100 & 0.062 \\ 0.092 & 0.077 & 0.100 & 0.100 & 0.092 \\ 0.123 & 0.107 & 0.107 & 0.092 & 0.138 \end{pmatrix}$$

$$A_{SAT} = \begin{pmatrix} 0.123 & 0.107 & 0.069 & 0.039 & 0.069 \\ 0.062 & 0.096 & 0.047 & 0.077 & 0.054 \\ 0.107 & 0.130 & 0.062 & 0.047 & 0.107 \\ 0.032 & 0.047 & 0.024 & 0.032 & 0.054 \\ 0.085 & 0.100 & 0.062 & 0.047 & 0.047 \\ 0.085 & 0.085 & 0.062 & 0.069 & 0.054 \\ 0.085 & 0.092 & 0.062 & 0.047 & 0.107 \end{pmatrix}$$

$$D_{UAV} = \begin{pmatrix} 1.000 & 0.168 & 0.016 & 0.077 & 0.092 & 0.153 & 0.107 \\ 0.047 & 1.000 & 0.107 & 0.123 & 0.062 & 0.123 & 0.016 \\ 0.092 & 0.077 & 1.000 & 0.062 & 0.107 & 0.123 & 0.123 \\ 0.077 & 0.198 & 0.077 & 1.000 & 0.047 & 0.153 & 0.062 \\ 0.032 & 0.123 & 0.047 & 0.092 & 1.000 & 0.123 & 0.047 \\ 0.062 & 0.138 & 0.062 & 0.062 & 0.092 & 1.000 & 0.092 \\ 0.168 & 0.047 & 0.153 & 0.123 & 0.123 & 0.123 & 1.000 \end{pmatrix}$$

$$D_{SAT} = \begin{pmatrix} 1.000 & 0.198 & 0.092 & 0.032 & 0.107 & 0.153 & 0.062 \\ 0.047 & 1.000 & 0.016 & 0.092 & 0.062 & 0.123 & 0.077 \\ 0.092 & 0.062 & 1.000 & 0.062 & 0.107 & 0.092 & 0.016 \\ 0.032 & 0.047 & 0.047 & 1.000 & 0.032 & 0.077 & 0.016 \\ 0.032 & 0.107 & 0.016 & 0.000 & 1.000 & 0.123 & 0.047 \\ 0.077 & 0.123 & 0.062 & 0.032 & 0.123 & 1.000 & 0.092 \\ 0.107 & 0.032 & 0.092 & 0.032 & 0.092 & 0.077 & 1.000 \end{pmatrix}$$

Table 8. Representative values of parameter Δ_k, $k = 1,...,7$ used in calculations.

$\Delta_k = 0$	No development in technology area *k*.
$\Delta_k = 1$	Same development in technology area *k* than in Table 5.
$\Delta_k = 1.1$	10 % more development in technology area *k* than in Table 5.
$\Delta_k = 1.5$	50 % more development in technology area *k* than Table 5.
$\Delta_k = 2$	100 % more development in technology area *k* than Table 5.
$\Delta_k = 0.9$	10 % less development in technology area *k* than in Table 5.
$\Delta_k = 0.5$	50 % less development in technology area *k* than in Table 5.

Appendix 2

Table 9. Statistical analysis of Figure 2 with full set of parameters is outlined in the table. On the left, the 10 year input data from questionnaires Q1 (%) and Q2 (Score) are shown. The confidence interval (lower, upper) is calculated at point x_0 = 1.5 to show the idea of calculating the 90 % confidence level curves in Figure 2. The Excel worksheet used in the calculations is available from (Zaiontz, 2015).

Confidence Intervals						
Score (x)	% (y)		Confidence interval for the forecasted value			
0.91	4.93					
1.11	7.13		n	6		
1.27	10.10		df	4		= n - 2
1.53	8.70		mean(x)	1.36		= AVERAGE(x)
1.59	10.07		x_0	1.5		
1.76	12.70		$\hat{y}_0$	9.979881		= FORECAST(y,x,x_0)
			s_{Res}	1.329159		= STEYX(y,x)
			SS_x	0.505552		= DEVSQ(x)
			se	0.600268		= s_{Res}*SQRT(1/n+(x_0-$\bar{x}$)^2/SS_x)
			t-crit	2.131847		= TINV(0.10,df)
			lower	8.700201		= $\hat{y}_0$ - t-crit * se
			upper	11.25956		= $\hat{y}_0$ + t-crit * se

III

Number of System Units Optimizing the Capability Requirements through Multiple System Capabilities

Vesa Kuikka

Journal of Applied Operational Research, 8(1), 2016, pp. 26-41

Available at http://www.orlabanalytics.ca/jaor/archive/v8/n1/jaorv8n1p26.pdf

Journal of Applied Operational Research (2016) Vol. **8**, No. 1, 26–41 ISSN 1735-8523 (Print), ISSN 1927-0089 (Online)

www.orlabanalytics.ca

Number of system units optimizing the capability requirements through multiple system capabilities

Vesa Kuikka [1,*]
[1] Finnish Defence Research Agency, Riihimäki, Finland

Received 09 May 2016
Accepted 11 August 2016

Keywords:
Benefit-cost analysis
Capability
Functionality
Optimal number of system units
Probabilistic modeling
Quantity discount
System of systems

Abstract—How to determine the optimal number of systems and system units is a common problem in many fields of business and public organizations. In this study we present a mathematical model for calculating the optimal number of systems to meet capability demand when a particular demand can be addressed by multiple distinct systems. In the model budgeted resource constraints, saturated and non-saturated capabilities and quantity discounts are considered. We present a numerical example for three capability areas and two systems but the model allows for any number of capability areas (or functionalities) and systems.

Published 01 December 2016

Introduction

In this section related work and concepts in the literature are discussed following with discussion of similar concepts in this paper. The outline of this paper is at the end of this section. Optimal resource allocation (Bridgeman, 2013) is vital in many fields of business and public planning. Ideas presented in this paper are demonstrated in a military context (Neaga *et al*., 2009; Kerr *et al*., 2007; Smith & Oosthuizen, 2011; Smith *et al*., 2012) but the same principles can be applied in many other fields (Webb, 2008). Resource reservation models in public health emergencies have been studied in (Bridgeman, 2013). The main research question is how to allocate the limited available budgeted funds to alternative systems and equipment. The optimal configuration of systems is determined by the maximum capability (Kudryavtsev *et al*., 2014; Kerr *et al*., 2007) of the final system of systems (Neaga *et al*., 2009; Smith & Oosthuizen, 2011; Smith *et al*., 2012; Sage & Cuppan, 2001; Biltgen 2007). In this respect the analysis can be cost-benefit analysis or cost-effectiveness analysis (Kee, 1999). In the literature, fair optimization (Golden, 2013; Denda *et al*., 2000; Ogryczak *et al*., 2014) and other resource allocation methods (Luss, 1999; Harris, 1913, Bourdon *et al*., 2014) have been studied extensively.

Capabilities-based planning is planning, under uncertainty, to provide capabilities suitable for a wide range of modern-day challenges and circumstances while working within an economic framework that necessitates choice (Davis, 2002). One definition of capability is the ability to execute a specified course of action (Biltgen, 2007; Biltgen & Mavris, 2006). To this end scenarios are an essential tool for describing different operations and missions. A model for the total capability composed of capability areas is in use in military forces in several countries (JETCD, 2009). Military capability areas are defined as statistically independent sets of functionalities. These functionalities have hierarchies that provide exact definitions of capability areas (JETCD, 2009; Smith *et al*., 2012). The same concept can be used for lower system level functionalities. Modeling can be conducted in various levels of granularity (Sage *et al*., 2001). It is a modeler's decision to design an adequate level of the model within the requirements of the problem (Smith *et al*., 2012).

Military capability is often defined as the amount of armament. Different classes of armament are difficult to quantify. Quantified Judgment Method (QJM) is a mathematical method quantifying weapons, ordnance, fighters etc. (Dupuy, 1979). The method is based on historical data and scoring different armaments. QJM is related with probabilistic definition of capability. System of systems is a common concept in system analysis and different mathematical models have been proposed (Golden, 2013). Multi-objective optimization is an area of multiple criteria decision making, that is concerned with mathematical optimization problems involving more than one objective function to be optimized simultaneously (Andersson, 2000; Weck, 2004). One of those methods is to model a system of systems as a network of serial and parallel units.

* *Correspondence*: Vesa Kuikka, Finnish Defence Research Agency,
PO BOX 10. Tykkikentäntie 1. FI-11311 Riihimäki, Finland
E-mail: vesa.kuikka@mil.fi

In this paper capability is defined as the probability of successful operation or mission. Our earlier papers (Kuikka & Suojanen, 2014; Kuikka *et al*., 2015a, 2015b; Suojanen *et al*., 2015) considered high level capability areas but in this paper we use the same method for functionalities. In this paper, a coherent way of constructing a model of systems of systems with desired level of granularity is presented. The concept of embedded systems describes our approach because new systems are connected to existing systems. The existing systems are divided in two sets of systems, parallel and serial, in relation to the new system functionalities.

Our earlier papers concerned technological forecasting in 1 and 10 years' time horizons. Because the focus of our work has been in technological forecasting, a straightforward and not too detailed approach is needed. Not enough information is available about technological characteristics of future systems and their possible use. Secondly, capability planning is only one part of modeling consisting of systems modeling and resource optimization. An approach is needed with the possibility to adjust the granularity level of modeling. This does not exclude the possibility of developing different probabilistic or deterministic models for capabilities. More detailed sub-models can be used in our modular framework. In fact, our goal is to provide a model for combining different models together.

In our earlier papers (Kuikka & Suojanen, 2014; Kuikka *et al*., 2015a, 2015b; Suojanen *et al*., 2015), we considered general concepts of systems like satellites and UAVs without taking into account how many satellites and UAVs actually are deployed in different scenarios and planning horizons. The main objective of this paper is to present a model for designing the optimal number of units. We extend the modeling method for multiple systems and unlimited number of functionalities. The same method is used for high level capability areas and system level functionalities. In this respect capability areas and system functionalities are commensurate concepts.

Basic probability theory is used in capability modeling in cases when there are multiple options of meeting particular set of requirements. Equations of not mutually exclusive events are applied for multiple systems capabilities, i.e. probabilities of successful execution of tasks, operations or missions (https://en.wikipedia.org/wiki/Mutual_exclusivity, accessed 05 May 2008). Two different domains of studying are identified: the number of units is saturated or non-saturated. In a saturated state the number of units exceeds a threshold value where the number of units is so high that more than one unit is available for the same operation. In other words, overlapping or redundant capability is available. Especially, in military context this is a desirable state of affairs. In many business situations it is important to have spare parts to avoid interruptions in production or other business functions. We examine the two domains of saturation and their relationship with the number of optimal units. These two domains provide a scope for planning reasonable allocations of units.

Another favorable result will be that proportional number of units is independent of the total budget of funds. As a consequence of this feature conclusions can be drawn without knowing the exact amount of funds available. We omit the technical details of not having money for a single unit or how to allocate resources in transition phases from one domain to another. We also examine the sensitivity of the optimal number of units to a model parameter. In some cases the model parameter can be used in calibrating the model if in a complex problem realistic modeling is difficult. This is an example of extending the theoretical model heuristically.

The outline of the paper is as follows. In the following two sections we present the capability model for two systems and optimal number of units for two systems. Next, some model calibration and sensitivity aspects are discussed. Then quantity discounts are modelled. Next, the general model for multiple capabilities, functionalities, systems and system units is described. This is then followed by a numerical example for three capability areas and two systems. This demonstrates how the results can be presented as ratios of the number of system units for an unknown budget and as actual number of system units for a known budget. Expert opinions from a questionnaire have been used as the input data for the numerical calculations. Conclusions summarize the results of the paper.

Two systems in parallel

In this section we deal with a particular capability demand that can be addressed by multiple distinct systems. This capability may consist of several system capabilities. We assume that the capability is available at present and two new systems are deployed with the existing systems. In the following equations, these three system capabilities are denoted K, X_1, X_2. The system capability can be produced by any one of the three alternative systems or a combination of them. If other system capabilities are necessary for the capability under consideration, a system denoted M is needed in the model.

In this section we present the model for the deployment of two new systems *1* and *2* in parallel with the existing system or systems k (Kuikka & Suojanen, 2014; Kuikka *et al*., 2015a). In addition, system m is in series with all these systems.

Systems *m* and *k* may be composed of several systems. Later, in this paper we present the general model where the number of systems is not limited. System of systems for two systems is described by the following equations:

$$p = MK \tag{1}$$
$$p + p_1 = M(X_1 + K - X_1K) \tag{2}$$
$$p + p_2 = M(X_2 + K - X_2K) \tag{3}$$
$$p + p_{12} = M(X_1 + X_2 + K - \alpha X_1X_2 - X_1K - X_2K + \alpha X_1X_2K) \tag{4}$$

In Equations (1 – 4) *p* is the initial capability value of the system of systems when systems *1* and *2* have not been deployed. Equation (1) describes the initial system of systems when systems *m* and *k* are in use with capability values *M* and *K* correspondingly. In Equation (1) system *k* has the same functionalities as systems *1* and *2*. The system capabilities of systems *1* and *2* are denoted X_1 and X_2. All the auxiliary and necessary functionalities are included in system *m*. Equations (2 – 4) describe system of systems with system *1*, system *2* and both systems *1* and *2* deployed. In Equations (2 – 4) p_1, p_2 and p_{12} are the additional capabilities provided by the new systems. The general principle of probability is that the serial systems' capabilities are multiplied and the parallel systems' capabilities are added with their common capabilities subtracted. This is the law of not mutually exclusive events. The parameter α describes possible joint effects of systems *1* and *2*.

In Equation (1) we assume that the systems *m* and *k* together explain the initial value of the capability. In our earlier papers we assumed that Equation (1) holds also for the future development of the capability values. The only restriction is that the product of *M* and *K* is *p*. In some cases the value of *K* has decreased as a result of deploying new systems in parallel with system *k* which has been used less because of other alternatives (Kuikka *et al*., 2015a). If Equation (1) does not hold, system *k* should be modeled in a more detailed level. This aspect will be discussed later in this paper.

We use as a concrete example satellites and UAVs. Equations (2) and (3) describe situations where only satellites (system *1*) or UAVs (system *2*) are deployed with systems *k* and *m* correspondingly. Equation (4) describes the situation where both satellites and UAVs are in use. Left hand sides of Equations (2-4) show that p_1, p_2 and p_{12} are changes to the initial value *p*. The values of p_1, p_2, p_{12} and *p* are input values for the model.

Equation (4) has the parameter α describing possible cooperative effects of systems *1* and *2*. Another need for the parameter may be that the functionalities of systems *1* and *2*, in fact, are the same but they have different time or spatial characteristics. The parameter may be regarded as artificial or undesirable in the modeling and a more detailed structure of sub-systems or a more exact definition of the probabilities as a function of time and spatial characteristics should be considered. Expressed with other words, the need for an additional parameter like α indicates that sub-systems may not be modeled in a proper way having no shared or common functionalities. In practice, modeling abstract systems and their functionalities may be a challenging task. In the basic model the value of $\alpha = 1$.

We ignore joint effects between other parallel systems but these can be considered in a similar manner. Later, we study also effects of the α parameter on the optimal number of the systems. Another motivation for the α parameter is sensitivity diagnostics of the model. Sensitivity to model parameters should be taken into account when the results are interpreted.

We get from Equations (2-4) with the help of Equation (1) Equations (5-7):

$$p_1 = MX_1 - X_1p \tag{5}$$
$$p_2 = MX_2 - X_2p \tag{6}$$
$$p_{12} = MX_1 + MX_2 - \alpha MX_1X_2 - X_1p - X_2p + \alpha X_1X_2p. \tag{7}$$

From Equations (5-7) we get for p_{12}

$$p_{12} = p_1 + p_2 - \alpha MX_1X_2 + \alpha X_1X_2p \tag{8}$$

and

$$p_{12} = p_1 + p_2 - \alpha p_1X_2. \tag{9}$$

Now we can solve the value of the system capability X_2

$$X_2 = \frac{p_1 + p_2 - p_{12}}{\alpha p_1}. \tag{10}$$

Similarly the value of the system capability X_1 is

$$X_1 = \frac{p_1 + p_2 - p_{12}}{\alpha p_2}. \tag{11}$$

From Equation (5) or (6) M can be solved

$$M = \frac{\alpha p_1 p_2}{p_1 + p_2 - p_{12}} + p \tag{12}$$

and from Equation (1) K can be solved

$$K = (1 + \frac{\alpha p_1 p_2}{p(p_1 + p_2 - p_{12})})^{-1}. \tag{13}$$

Optimal number of units of two systems

Usually there are constraints for the number of systems which may be due to budget, time or personnel. Here, we consider the most common limitation, the amount of funds. In Equation (14) the amount of money available, denoted by C, is used for purchasing a number n_1 of system *1* and a number of n_2 of system *2*. Functions g_i describe the cost of systems $i=1, 2$.

$$C = g_1(n_1) + g_2(n_2) \tag{14}$$

We use the method of Lagrange multipliers for the optimization problem. First, we maximize the value of a capability area with two systems. Generalizations for multiple systems and for more than one capability area (or functionality) are presented later in this paper. The Lagrangian function L for two systems with the Lagrangian multiplier λ is the following:

$$L(n_1, n_2, \lambda) = P(n_1, n_2) + \lambda(g_1(n_1) + g_2(n_2) - C) \tag{15}$$

where the capability value P for the total capability value is

$$P(n_1, n_2) = (M - p)A(n_1, n_2) + p, \tag{16}$$

where

$$A(n_1, n_2) = 1 - (1 - x_1)^{n_1}(1 - x_2)^{n_2}. \tag{17}$$

In Equation (19) x_1 and x_2 are capabilities of one of systems *1* and one of systems *2* correspondingly, for example, one satellite and one UAV. We have the identity $1 - X_i = (1 - x_i)^{n_i}, i = 1,2$. Partial derivatives are

$$\frac{\partial L}{\partial n_1} = -ln(1 - x_1)(1 - x_1)^{n_1}(1 - x_2)^{n_2} + \lambda g'_1(n_1) = 0, \tag{18}$$

$$\frac{\partial L}{\partial n_2} = -ln(1 - x_2)(1 - x_1)^{n_1}(1 - x_2)^{n_2} + \lambda g'_2(n_2) = 0 \text{ and} \tag{19}$$

$$\frac{\partial L}{\partial \lambda} = g_1(n_1) + g_2(n_2) - C = 0. \tag{20}$$

It follows from Equations (18) and (19)

$$\frac{g'_1(n_1)}{g'_2(n_2)} = \frac{ln(1 - x_1)}{ln(1 - x_2)} = \frac{n_2}{n_1}\frac{ln(1 - X_1)}{ln(1 - X_2)} \tag{21}$$

and the partial derivative with respect to λ gives Equation(14).

Next, we present a more general model for the parallel systems $A(n_1, n_2)$. If $n_1 \le n'_1$ system capabilities in X_1 are additive and if $n_2 \le n'_2$ system capabilities in X_2 are additive. This provides a model for the situation where individual systems act in distinct areas or time periods. For example, ten satellites orbiting the earth with only one satellite at a time being capable of observing the area of operation. If the number of satellites exceeds n'_1 redundancy exists because more than one satellite at a time may be capable of making observations. Similar consideration can be made concerning UAVs. The actual value of n'_2 for UAVs depends on the size of the area of operation. The phenomenon may be compared with saturation in natural sciences. We present two alternatives, saturated and non-saturated, models for the dependency between the system capability and the number of units. The model is described in Equations (22-25).

$$A(n_1, n_2) = 1 - (1 - n'_1 x_1)(1 - n'_2 x_2),\ n_1 \le n', n_2 \le n'_2 \tag{22}$$

$$A(n_1, n_2) = 1 - (1 - n'_1 x_1)(1 - x_1)^{n_1 - n'_1}(1 - n_2 x_2),\ n_1 > n'_1, n_2 \le n'_2 \tag{23}$$

$$A(n_1, n_2) = 1 - (1 - n_1 x_1)(1 - n'_2 x_2)(1 - x_2)^{n_2 - n'_2},\ n_1 \le n'_1, n_2 > n'_2 \tag{24}$$

$$A(n_1, n_2) = 1 - (1 - n'_1 x_1)(1 - x_1)^{n_1 - n'_1}(1 - n'_2 x_2)(1 - x_2)^{n_1 - n'_2},\ n_1 > n'_1, n_2 > n'_2 \tag{25}$$

Equation (21) is for $n_1 > n'_1$ and $n_2 > n'_2$. It is a straight forward calculation to show that $\ln(1 - x_1)$ is replaced by $x_1(1 - n_2 x_2)$ if $n_1 \le n'_1$ and $\ln(1 - x_2)$ is replaced by $x_2(1 - n_1 x_1)$ if $n_2 \le n'_2$. For example, if $n_1 \le n'_1$ and $n_2 \le n'_2$

$$\frac{g'_1(n_1)}{g'_2(n_2)} = \frac{x_1(1 - n_2 x_2)}{x_2(1 - n_1 x_1)} = \frac{n_2}{n_1}\frac{X_1(1 - X_2)}{X_2(1 - X_1)}, \tag{26}$$

where $X_i = n_i x_i,\ i = 1,2$. We need to specify the functional forms of g_1 and g_2. The simplest choice is

$$g_i(n_i) = h_i n_i,\ i = 1,2, \tag{27}$$

where h_i are constant prices of one system unit $i=1,2$.

From Equations (26) and (27) we get

$$\frac{h_1}{h_2} = \frac{n_2}{n_1}\frac{X_1(1 - X_2)}{X_2(1 - X_1)}.$$

In the questionnaire results the average values of X_1 and X_2 are 0.396 and 0.599 correspondingly. If the value of h_1/h_2 is 10 we get the numerical value:

$$\frac{n_2}{n_1} = \frac{h_1}{h_2}\frac{X_2(1 - X_1)}{X_1(1 - X_2)} = 22.7 \approx 23.$$

If $n_1 > n'_1$ and $n_2 > n'_2$ we get

$$\frac{n_2}{n_2} = \frac{h_1}{h_2}\frac{ln(1 - X_2)}{ln(1 - X_1)} = 18.1 \approx 18.$$

In the first case both systems of satellites (system *1*) and UAVs (system *2*) are not saturated. In the second case both systems are saturated. These two results give us the first simple rules of thumb about the ratio of the number of UAVs to the number of satellites. The result is that in the non-saturated state a higher number of UAVs than 23 should be procured prior to the first satellite. If $n_1 \le n'_1$ and $n_2 > n'_2$ we get $n_2 / n_1 = 13.9 \approx 14$ and if $n_1 > n'_1$ and $n_2 \le n'_2$ we get $n_2 / n_1 = 29.6 \approx 30$. The first case is more likely for a low budget country. In this case more than 14 UAVs should be procured before the first satellite. The second case may be possible for a scenario not requiring continuing satellite capabilities yet needing a high number of UAVs to cover spatial and timing requirements. For example, if $n'_1 = 0$ or $n'_1 = 1$ and $n_1 = 1$, more than 60 UAVs should be procured prior to the second satellite. More precisely, more than

$45\ (=30+30/2)$ UAVs should be procured prior to the second satellite if we consider average ratios of UAV and satellite units. The number 45 UAVs is the minimum number and there may a need for more profound operational analysis to find out the best alternative between 45 and the maximum number $75\ (=60+30/2)$ UAVs. These examples show how the theory provides an interval and room is left for more detailed analysis. The reason for the big difference between minimum and maximum numbers is the assumption of one satellite paying 10 times more than one UAV. In practice, the actual market prices of the systems $h_i,\ i=1,2$ or the corresponding ratio of the prices should be used in the equations.

Model calibration

In this section we study the effect the parameter α introduced in Equation (4). Sensitivity to modeling is one important aspect to consider and understanding the limitations of models is essential when models are applied in real world problems. Another possibility is to adjust the model to empirical data or to a theoretical sub-model that better describe the relationships between capabilities and unit quantities.

When α can take values other than one $A(n_1,n_2)$ can be expressed when $n_1 > n'_1$ and $n_2 > n'_2$ as

$$A(n_1,n_2)=X_1+X_2-\alpha X_1X_2=1-(1-x_1)^{n_1}+1-(1-x_2)^{n_2}-\alpha\left(1-(1-x_1)^{n_1}\right)\left(1-(1-x_2)^{n_2}\right)=$$
$$2-\alpha+(\alpha-1)\left((1-x_1)^{n_1}+(1-x_2)^{n_2}\right)-\alpha(1-x_1)^{n_1}(1-x_2)^{n_2}.$$

Because of our probabilistic interpretation of capabilities, $A(n_1,n_2)$ takes values between 0 and 1. This sets physical limits for α also. The partial derivatives of $P(n_1,n_2)$ are

$$\frac{\partial P(n_1,n_2)}{\partial n_1}=(M-p)\frac{\partial A(n_1,n_2)}{\partial n_1}=(M-p)\ln(1-x_1)(1-x_1)^{n_1}\left\{\alpha\left(1-(1-x_2)^{n_2}\right)-1\right\}=$$
$$(M-p)\ln(1-X_1)(1-X_1)(\alpha X_2-1) \text{ and}$$
$$\frac{\partial P(n_1,n_2)}{\partial n_2}=(M-p)\frac{\partial A(n_1,n_2)}{\partial n_2}=(M-p)\ln(1-x_2)(1-x_2)^{n_2}\left\{\alpha\left(1-(1-x_1)^{n_1}\right)-1\right\}=$$
$$(M-p)\ln(1-X_2)(1-X_2)(\alpha X_1-1).$$

We get as in Equation (21)

$$\frac{g'_1(n_1)}{g'_2(n_2)}=\frac{n_2}{n_1}\frac{\ln(1-X_1)}{\ln(1-X_2)}\frac{(1-X_1)(\alpha X_2-1)}{(1-X_2)(\alpha X_1-1)},\ n_1>n'_1 \text{ and } n_2>n'_2. \tag{28}$$

If $n_1 \le n'_1$ and $n_2 \le n'_2$

$$\frac{\partial A(n_1,n_2)}{\partial n_1}=\frac{X_1}{n_1}(1-\alpha X_2) \text{ and}$$
$$\frac{\partial A(n_1,n_2)}{\partial n_2}=\frac{X_2}{n_2}(1-\alpha X_1).$$

We get as in Equation (26)

$$\frac{g'_1(n_1)}{g'_2(n_2)}=\frac{n_2}{n_1}\frac{X_1(1-\alpha X_2)}{X_2(1-\alpha X_1)},\ n_1\le n'_1 \text{ and } n_2\le n'_2. \tag{29}$$

If $n_1 > n'_1$ and $n_2 \le n'_2$

$$\frac{g'_1(n_1)}{g'_2(n_2)}=\frac{n_2}{n_1}\frac{\ln(1-X_1)(1-X_1)(1-\alpha X_2)}{X_2(\alpha X_1-1)} \text{ and} \tag{30}$$

if $n_1 \le n'_1$ and $n_2 > n'_2$

$$\frac{g'_1(n_1)}{g'_2(n_2)} = \frac{n_2}{n_1}\frac{ln(1-X_2)(1-X_2)(1-\alpha X_1)}{X_1(\alpha X_2 - 1)}. \tag{31}$$

We use notation $Z_i(\alpha), i = 1,2$ in Equations (28-31) and the price functions of Equation (27):

$$\frac{g'_1(n_1)}{g'_2(n_2)} = \frac{h_1}{h_2} = \frac{n_2}{n_1}\frac{Z_2(\alpha)}{Z_1(\alpha)}, \tag{32}$$

where

$Z_i = \frac{ln(1-X_i)(1-X_i)}{(\alpha X_i - 1)}$ if $n_i \geq n'_i$ and $Z_i = \frac{X_i}{1-\alpha X_i}$ if $n_i < n'_i$,$i = 1,2$.

Figure 1 shows the ratio n_2/n_1 as a function of α in the four cases with systems *1* and *2* in saturated and non-saturated states. The linear function in Equation (27) for the prices is used with the assumption h_2/h_1=10, i.e. one satellite is 10 times more expensive than one UAV. In the next section we demonstrate a different function for the system unit prices $g_i(n_i)$ for system *i*. Again, a realistic function or actual empirical data sets should be used. The basic theory has the value of $\alpha = 1$.

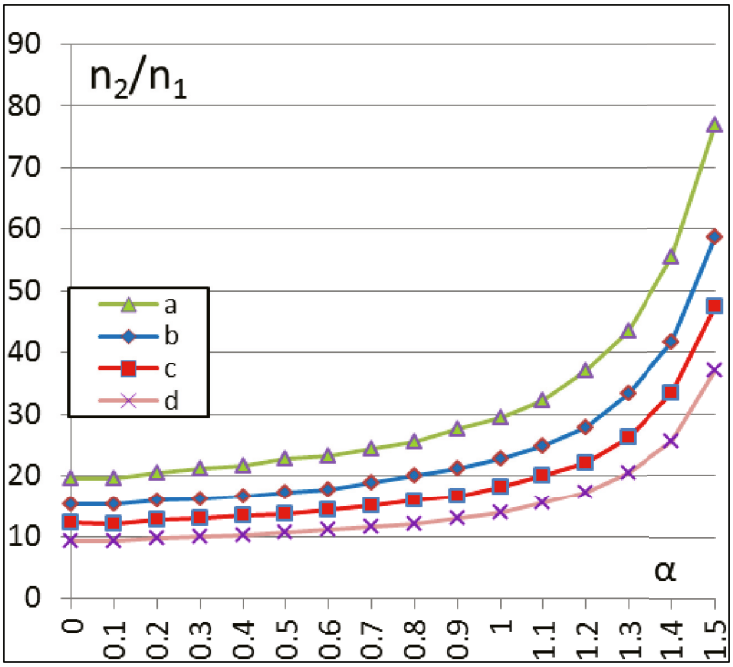

Figure 1. Ratio of the number of UAVs to the number of satellites as a function of α for the four cases of two systems in saturated or non-saturated states: $a) n_1 \geq n'_1, n_2 < n'_2, b) n_1 < n'_1, n_2 < n'_2$, $c) n_1 \geq n'_1, n_2 \geq n'_2$ and $d) n_1 < n'_1, n_2 \geq n'_2$.

In practice, systems integration is technically difficult and no full power of systems can be achieved. This may result in different systems and operating personnel working separately. This can be a reason to calibrate the model and use a parameter value less than one: $\alpha < 1$. As we can see form Figure 1, the model is not very sensitive to the parameter values of $0 \leq \alpha < 1$. In some rare occasions $\alpha < 0$ when the systems operate jointly and generate more capability than the sum of individual systems have. This may be a sign for a need for more detailed modeling or more precise definitions of functionalities and capabilities.

Another consideration is that, e.g. in case of sensors, the trust that the operators have in particular systems could influence the decisions. In some cases the information from some of the sensors would not be accepted without corroboration of other sensors.

In many cases, procurement process proceeds incrementally. New systems are deployed in the span of years or tens of years. Budget constraints, infrastructure development and personnel training introduce limitations. Model structure, capability values and model parameters may change during the course of time when new systems are deployed and operation of existing systems becomes more efficiently.

Model for quantity discount

Next we consider a more realistic model for quantity discounts. In previous sections a simple linear function of Equation (27) was used. Usually quantity discounts are used when systems are procured. Also, when systems are built in house, planning and manufacturing first system units is more costly than systems manufactured after the initial stages of production. Here we use the model in Equations (33) to demonstrate our general method. However, in modeling actual discount function or numerical realistic data should be used. The discount is suppressed for higher number of system units with the parameter $\delta_i, i = 1,2$. Figure 2 shows quantity discounts in percentages as a function of δ_i for some representative values of δ_i.

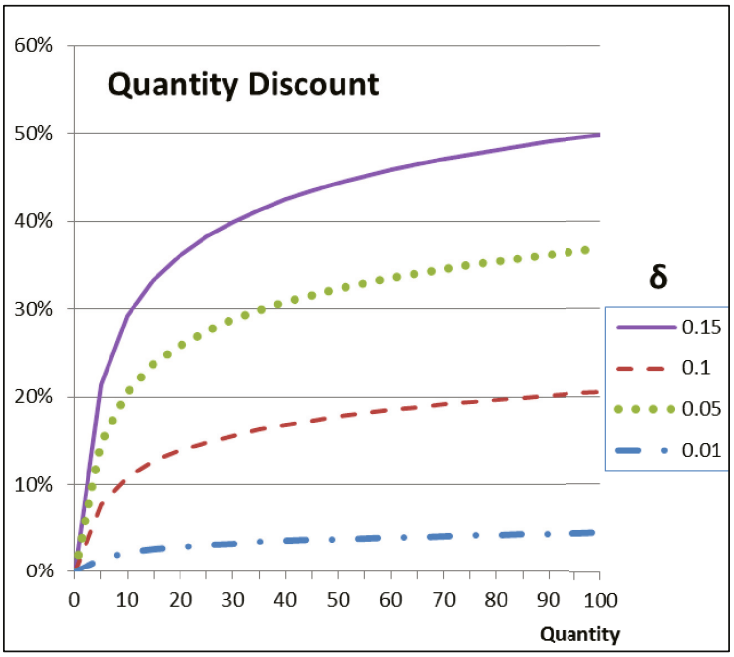

Figure 2. Quantity discount in percentages from Equation (33).

System unit prices are

$$g_1(n_1) = h_1 n_1^{1-\delta_1} \text{ and} \qquad (33)$$
$$g_2(n_2) = h_2 n_2^{1-\delta_2},$$

where δ_1 and δ_2 are constants. Equations (33) reduce to Equation (27) in the case that $\delta_1 = 0$ and $\delta_2 = 0$. We get

$$\frac{g'_1(n_1)}{g'_2(n_2)} = \frac{h_1}{h_2} \frac{(1-\delta_1) n_1^{-\delta_1}}{(1-\delta_2) n_2^{-\delta_2}}.$$

If $n_1 > n'_1$ and $n_2 > n'_2$

$$\frac{h_1}{h_2}\frac{(1-\delta_1)n_1^{-\delta_1}}{(1-\delta_2)n_2^{-\delta_2}}=\frac{n_2}{n_1}\frac{\ln(1-X_1)}{\ln(1-X_2)}.$$

And for the four cases of two systems in saturated or non-saturated states we have

$$\frac{n_2^{1-\delta_2}}{n_1^{1-\delta_1}}=\frac{h_1}{h_2}\frac{(1-\delta_1)}{(1-\delta_2)}\frac{Z_2}{Z_1}. \tag{34}$$

The ratios of the number of UAV units to the number of satellite units with numerical values X_1=0.39, X_2=0.60 and $\delta_1=\delta_2$ are shown in Figure 3. If $\delta_1=2\delta_2$, for example, curves would be almost flat compared to Figure 3. This may be a more realistic discount policy making the pricing more neutral between satellites and UAVs.

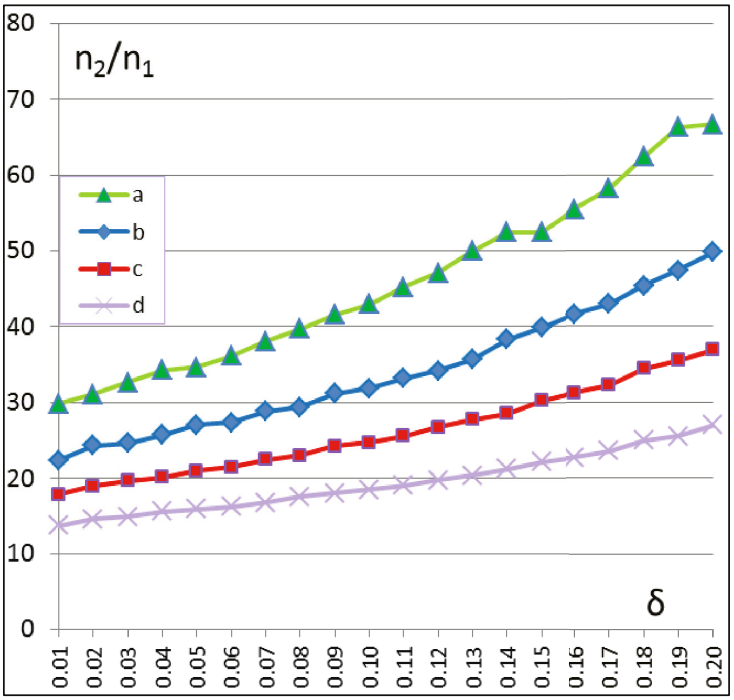

Figure 3. Ratio of the number of UAVs to the number of satellites as a function of δ for the four cases of two systems in saturated or non-saturated states: $a)\, n_1 \geq n'_1, n_2 < n'_2, b)\, n_1 < n'_1, n_2 < n'_2$, $c)\, n_1 \geq n'_1, n_2 \geq n'_2$ and $d)\, n_1 < n'_1, n_2 \geq n'_2$.

General model for system of systems and their functionalities

In this section we describe the idea of modeling multiple systems. The systems may have parallel or serial functionalities. As our definition of capability is the probability of a successful mission or task, the necessary functionalities appear as serial sub-systems and the alternative functionalities appear as parallel sub-systems. In the general procedure the systems are decomposed into functionalities. The granularity of the model depends on the requirements of the model and the fidelity of the input data. As mentioned before, special attention may be given to possible time and spatial characteristics of the model. An example of three systems consisting of different number of sub-systems (or functionalities) is show in Figure 4. The first system has three sub-systems, one of which is common for all three systems (oval), one is common for two systems (triangle) and one is not included in other two systems (diamond).

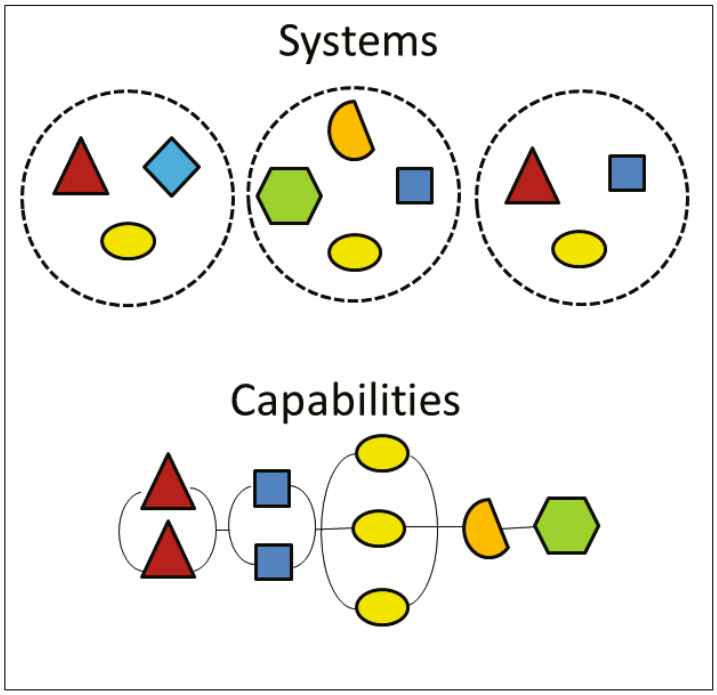

Figure 4. Three systems consisting of six different functionalities (10 sub-systems). One functionality has no effect on the capability under investigation.

In the lower part of Figure 4 the systems are disassembled and shown as serial and parallel sub-systems. Equations describing the situation can now be written down easily for one, two and three parallel systems and multiple serial systems. The only difficulty may arise with more than two parallel systems (sub-systems). This can be overcome with an iterative procedure. We can add a new system to the existing configuration by considering systems *1* and *2* as one system of systems and using the equations iteratively. The same principle can be used when the input data is collected from expert questionnaires or from simulation models. Input values p_1, p_2 and p_{12} are complemented with p_{123} values where all the systems *1*, *2*, *3* with systems *m* and *k* are in use. System *3* is a new system not included in *k*.

If system *3* is included in *k* the equations should be modified accordingly. This is a case when a more detailed modeling is needed for the system *k*. Now, the system capability *K* is expressed with X_3 and *K'* as

$$K = 1 - (1 - X_3)(1 - K').$$

K' is the system remaining after *3* has been removed and modeled individually. For collecting the input data the values of p_3 are obtained from expert opinions or simulated. The value of *K'* can be solved because the value of *K* is known from Equation (13). After a number of modeling phases *K* consists only of one system and no general parallel system of systems *K* exists in the final model.

The general form for total capability value *P* is

$$p + p_{1,\ldots,I} = \prod_{f=1,F} \left((M_f - p_f) A_f + p_f \right) = \prod_{f=1,F} \left(M_f A_f + (1 - A_f) p_f \right),$$

where *f* is the functionality and p_f are the initial values of functionalities *f* and

$$A_f = 1 - \prod_{i=1,I} (1 - X_{if})$$

for *I* systems. Some infrastructure capabilities may be suitable for modeling as a serial sub-system. Energy or communications are possible candidates for M_f. In addition, if all except one germane parallel sub-system have been included in the model, K_f is regarded as a comparable sub-system with others which can be evaluated with the same method as other sub-systems in the model.

General model for the number of units

In this section we present a general model for I systems and F functionalities (or capability areas). In the general case, the Lagrangian function takes the following form

$$L = \prod_{f=1,F}\left((M_f - p_f)A_f(n_1,n_2,...,n_I)\right) + p_f) + \lambda(C - \sum_{i=1,I} g_i(n_i)), \tag{35}$$

where

$$A_f(n_1,n_2,...,n_I) = 1 - \prod_{i=1,I}(1 - X_{if}), \tag{36}$$

and

$$X_{if} = 1 - (1 - x_{if})^{n_i}, \text{ if } n_i \geq n'_i,$$

$$X_{if} = n_i x_{if}, \text{ if } n_i < n'_i.$$

$X_{if} = 0$, if the functionality f does not exist in system i. The optimal number of system units can be calculated with the Lagrangian method as follows:

$$\frac{\partial L}{\partial n_i} = \sum_f (M_f - p_f)\frac{\partial A_f}{\partial n_i} - \lambda g'_i(n_i) = 0$$

and

$$\frac{\partial L}{\partial \lambda} = C - \sum_{i=1,I} g_i(n_i) = 0,$$

where

$$\frac{\partial A_f}{\partial n_i} = -\ln(1 - x_{if})(1 - x_{if})^{n_i}\prod_{\substack{i=1,I \\ i \neq n_i}}(1 - X_{if}) - \lambda g'_i(n_i) = 0, \text{ if } n_i \geq n'_i$$

and

$$\frac{\partial A_f}{\partial n_i} = -x_{if}\prod_{\substack{i=1,I \\ i \neq n_i}}(1 - X_{if}) - \lambda g'_i(n_i), \text{ if } n_i < n'_i.$$

These can be expressed in the following form:

$$\frac{\partial A_f}{\partial n_i} = -ln(1 - x_{if})A_f - \lambda g'_i(n_i) = 0, \text{ if } n_i \geq n'_i$$

and

$$\frac{\partial A_f}{\partial n_i} = \frac{x_{if}}{1 - n_i x_{if}} A_f - \lambda g'_i(n_i) = 0, \text{ if } n_i < n'_i.$$

A numerical example for three capability areas and two systems

Our numerical example shows that quantitative results can be achieved by modelling different aspects of capability optimization and combining the results in one model. For three functionalities (or capability areas) $f = 1,2,3$ and two systems $i = 1,2$ we get, if $n_i \geq n'_i$, i = 1, 2.

$$(M_1 - P_1)(1 - A_1)\ln(1 - x_{11}) + (M_2 - P_2)(1 - A_2)\ln(1 - x_{12}) + (M_3 - P_3)(1 - A_3)\ln(1 - x_{13}) + \lambda g'_1(n_1) = 0$$

$$(M_1 - P_1)(1 - A_1)\ln(1 - x_{21}) + (M_2 - P_2)(1 - A_2)\ln(1 - x_{22}) + (M_3 - P_3)(1 - A_3)\ln(1 - x_{23}) + \lambda g'_2(n_2) = 0.$$

Because

$\ln(1 - x_{if}) = \dfrac{\ln(1 - X_{if})}{n_i}$, i = 1,2 and f=1,2,3 we get

$$\frac{g'_2(n_2)}{g'_1(n_1)}\frac{n_2}{n_1}=\frac{\sum_{f=1,2,3}(M_f-P_f)(1-A_f)ln(1-X_{2f})}{\sum_{f=1,2,3}(M_f-P_f)(1-A_f)ln(1-X_{1f})}. \tag{37}$$

In Equation (37) we assumed that systems are in a saturated state in all the functionalities. In the general form for I systems and F functionalities we have

$$\frac{g'_{i_1}(n_{i_1})n_{i_1}}{g'_{i_2}(n_{i_2})n_{1_2}}=\frac{\sum_{f=1,F}(M_f-P_f)(1-A_f)Z_{i_1f}}{\sum_{f=1,F}(M_f-P_f)(1-A_f)Z_{i_2f}},\ i_1, i_2=1,\ldots,I, \tag{38}$$

where

$$Z_{if}=-\ln(1-X_{if})\text{ if } n_i\geq n'_i \text{ and } Z_{if}=\frac{X_{if}}{1-X_{if}}\text{ if } n_i<n'_i. \tag{39}$$

In Figures 5 and 6 we show saturated and non-saturated quantities with some representative values of α. As can be seen the saturated and non-saturated models give different values with high values of the system capability. Using α provides a method for calibrating the model. In our numerical examples we set α =1 because we have no special information for model calibration.

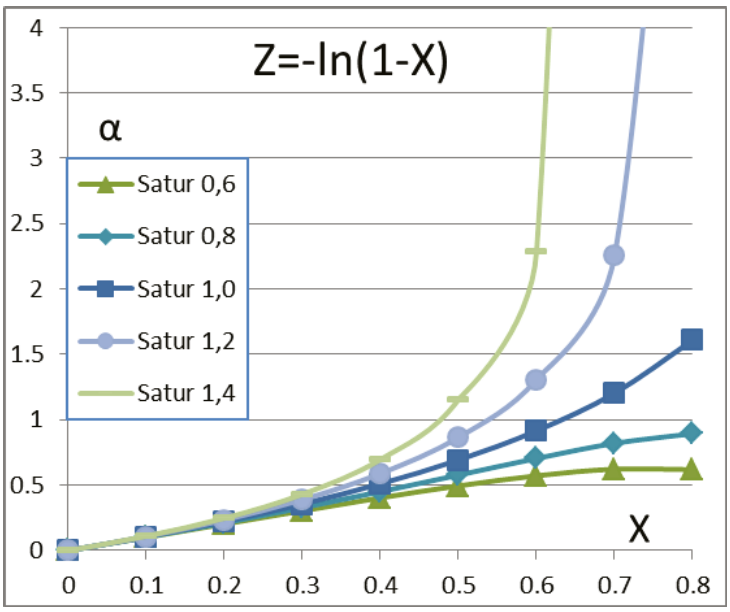

Figure 5. Quantity Z in saturated states for representative values of α as a function of system capability.

Scenarios are dealt with separately and consequently the same systems in functionalities can be in a saturated state in one scenario and in a non-saturated state in another scenario. In Equation (38) the quantity Z should be used accordingly. These variations of the theory expand the usability of the theory considerably.

We denote

$$R_i=\sum_{f=1,F}(M_f-P_f)(1-A_f)Z_{i,f}.$$

Again, we use the linear price function of Equation (27). A more general function or empirical data can be used as explained earlier in this paper. Ratio of system units can be expressed as

$$\frac{n_{i_1}}{n_{1_2}}=\frac{h_{i_1}}{h_{i_2}}\frac{R_{i_1}}{R_{i_2}}.$$

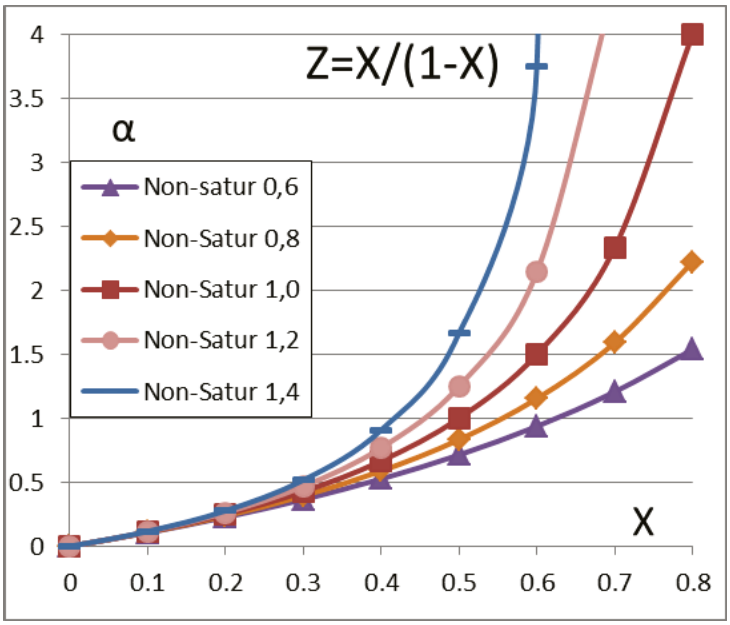

Figure 6. Quantity Z in non-saturated states for representative values of α as a function of system capability.

Next, we show how to use Equation (38) in practice. In Table 1 we reproduce the questionnaire data from our earlier work (Kuikka *et al*., 2015a). Expert opinions and evaluations for capability values have been enquired for three capability areas, three scenarios, two systems and two time horizons. Due to a small number of participants (5 military technology experts and 5 operational analysts) in the questionnaire the initial values have been provided in advance for the three capability areas and three scenarios. In Table 1 the rows for satellite and UAV system capability values X_{1f} and X_{2f} are from the questionnaire. In Table 1 we have all the necessary information for numerical calculations.

Numerical values of R_i for three scenarios and two systems for saturated and non-saturated states are given in Table 2. The values can be used in Equation (38) with different quantity discount relationships and possibly with mixed saturated and non-saturated states. In Table 3 the results are shown for linear price functions of Equation (27) and for saturated and non-saturated states.

Table 1. Values of system capabilities for three capability areas (f=1,2,3) and three scenarios. The initial values p_f for the three capability areas for each scenario are also shown (Kuikka *et al*., 2015a).

Capa-bility	*Years*	p_1			p_2			p_3		
		0.7	*0.5*	*0.5*	*0.8*	*0.6*	*0.4*	*0.9*	*0.3*	*0.4*
		Protection (f=1)			*Awareness (f=2)*			*Engagement (f=3)*		
		Sce1	*Sce2*	*Sce3*	*Sce1*	*Sce2*	*Sce3*	*Sce1*	*Sce2*	*Sce3*
X_{1f}	*1 y*	0.190	0.171	0.553	0.361	0.426	0.438	0.167	0.517	0.227
	10 y	0.132	0.290	0.765	0.518	0.447	0.705	0.625	0.423	0.182
X_{2f}	*1 y*	0.267	0.125	0.700	0.667	0.725	0.549	0.500	0.939	0.556
	10 y	0.182	0.354	0.844	0.698	0.663	0.777	0.938	0.740	0.556
M_f	*1 y*	0.857	0.780	0.561	0.892	0.694	0.562	0.912	0.364	0.479
	10 y	1.118	0.783	0.601	0,922	0.786	0.623	0.926	0.527	0.598
A_f	*1 y*	0.406	0.275	0.866	0.787	0.842	0.747	0.583	0.971	0.657
	10 y	0.289	0.541	0.963	0.855	0.813	0.934	0.977	0.850	0.636

Table 2. Values of R_i for saturated ($n_i \geq n'_i$) and non-saturated ($n_i < n'_i$) states.

R_i			*Scenario 1*	*Scenario 2*	*Scenario 3*
SAT	$n_1 \geq n'_1$	1 Year	0.0293	0.0477	0.0372
		10 Years	0,0556	0,0838	0,0379
UAV	$n_2 \geq n'_2$	1 Year	0.0540	0.0515	0.0645
		10 Years	0,0825	0,1405	0,0875
SAT	$n_1 < n'_1$	1 Year	0.0339	0.0549	0.0500
		10 Years	0,0652	0,1061	0,0634
UAV	$n_2 < n'_2$	1 Year	0.0782	0.0967	0.1029
		10 Years	0,1161	0,2365	0,1618

Table 3. Ratio of the number of UAVs to the number of satellites in saturated and non-saturated states.

		Scenario 1	*Scenario 2*	*Scenario 3*
$n_1 \geq n'_1$ and $n_2 \geq n'_2$	1 Year	18	11	17
	10 Years	15	17	23
$n_1 < n'_1$ and $n_2 < n'_2$	1 Year	23	18	21
	10 Years	18	22	26
$n_1 < n'_1$ and $n_2 \geq n'_2$	1 Year	16	9	13
	10 Years	13	13	14
$n_1 \geq n'_1$ and $n_2 < n'_2$	1 Year	27	20	28
	10 Years	21	28	43

Information in Table 3 can be used in various planning situations and the concrete quantities of units of systems make the interpretation of the results very easy. We give a couple of examples. UAVs are important in Scenario 3 in 10 years and especially in situations where satellites are in a saturated state and UAVs are not. In Scenario 1 the ratio of the number of UAVs to the number of satellites is decreasing in 10 years.

Conclusions

In this study we present a general modeling framework for optimizing capability of functionalities in system of systems used in operations or missions. We present a mathematical model for calculating the optimal number of system units to meet capability demand when a particular demand can be addressed by multiple distinct systems. The approach is cost-benefit analysis rather than cost-effectiveness analysis (Kee, 1999). The concepts can be applied in many fields of military, economy, business, and public life.

We generalize our earlier model (Kuikka & Suojanen, 2014; Kuikka *et al*., 2015a) for two systems and three capability areas for a desired configuration of capabilities (or functionalities), systems and number of system units. In modeling the relationship between a capability and the number system units we present a basic probabilistic model which can be in a saturated or in a non-saturated state. These alternatives with an option to adjust the model with phenomenological parameters provide a number of ways to apply the modelling method.

We demonstrate the model with questionnaire data of expert opinions in the context of military capability areas. Deployment of new satellite and UAV systems are studied in technology forecasting perspective. The optimal ratios of the number of UAVs to the number of satellites are calculated for 1 and 10 years forecasting horizons. The numerical example

shows how to calculate optimal number of satellites and UAVs to get maximal total capability with a budget constraint in three different scenarios. The eventual funds available determine the actual number UAVs and satellites to be acquired. We also show how to incorporate a model for quantity discounts as a mathematical formula or as an empirical data set.

Our modeling method is modular as more detailed sub-models or empirical data can be used for different aspects of the model: detailed system structures, different capability definitions, levels of granularity in systems and functionalities, more realistic relationships between the capabilities and the number of units, etc. One may use detailed modeling or simulation for detecting the relationships of particular concepts. This is possible because all the variables have a quantified empirical content. Capability, for example, is defined as the probability of successful operation or mission. In this respect our model can be taken as metamodeling for various future studies, where different modules of the theory can be evaluated or calculated with different methods, e.g. simulations, empirical data, more detailed technical models or expert opinions. Often quantity discounts are very specific and this part of the model can be exchanged with more appropriate information.

References

Andersson, J. (2000). A Survey of Multiobjective Optimization in Engineering Design, Technical Report: LiTH-IKP-R-1097.

Biltgen, P. T. (2007). A Methodology for Capability-Based Technology Evaluation for Systems-of-Systems, School of Aerospace Engineering Georgia Institute of Technology May 2007, PhD Thesis.

Biltgen, P. T., Mavris, D. N. (2006). A Methodology for Technology Evaluation and Capability Tradeoff for Complex System Architectures, 25th International Congress of the Aeronautical Science.

Bordon, S., Arbour, B., MacLeod, M. R. (2014). A method for hierarchically prioritizing capabilities with an application to military manned and unmanned aerial vehicles, Journal of Applied Operational Research 6(1), 39-47.

Bridgeman, D. (2013). Optimization Models for Resource Allocation Under Perturbation, Electronic Theses and Dissertations. Paper 151.

Davis, P. K. (2002). Analytic Architecture for CapabilitiesBased Planning, Mission-System Analysis, and Transformation. MR-1513-OSD, RAND National Defence Research Institute.

de Weck, O. L. (2004). Multiobjective Optimization: History and Promise, The Third China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical Systems, October 30-November 2, 2004, Kanazawa, Japan.

Denda, R., Banchs, A., Effelsberg, W. (2000). The Fairness Challenge in Computer Networks, J. Crowcroft, J. Roberts, Smirnov, M. (Eds.): QofIS 2000, LNCS 1922, pp. 208-220, Springer-Verlag Berlin Heidelberg.

Dupuy, T. N. (1979). Numbers, Predictions and War: Using History to Evaluate Combat Factors and Predict the Outcome of Battles, Bobbs-Merrill.

Golden, B. (2013). A Unified Formalism for Complex Systems Architecture, École Polytechnique, PhD Thesis.

Harris, F. W. (1913). How Many Parts to Make at Once, Factory, The Magazine of Management, Vol 10, No 2, 135-136.

J7 Joint Experimentation, Transformation, and Concepts Division (JETCD) (2009). Joint Capability Areas.

Kee, J. E. (1999). At What Price? Benefit-Cost Analysis and Cost-Effectiveness Analysis in Program Evaluation, The Evaluation Exchange, Volume V, Number 2/3.

Kerr, C., Phaal, R., Probert, D. (2007). A Framework for Strategic Military Capabilities in Defense Transformation, 11th International Command and Control Research and Technology Symposium 11th ICCRTS.

Kudryavtsev, D., Grigoriev, L. and Bobrikov, S. (2014). Strategy-focused and Value-oriented Capabilities: Methodology for Linking Capabilities with Goals and Measures, CEUR Workshop Proceedings, 1182:15-26.

Kuikka, V., Nikkarila, J-P., Suojanen, M. (2015a). A Technology Forecasting Method for Capabilities of a System of Systems, PICMET Conference.

Kuikka, V., Nikkarila, J-P., Suojanen, M. (2015b). Dependency of Military Capabilities on Technological Development, Journal of Military Studies, Vol 6, No 2.

Kuikka, V., Suojanen, M. (2014). Modeling the Impact of Technologies and Systems on Military Capabilities, Journal of Battlefield Technology, Vol. 17, No. 2, 9-16.

Luss, H. (1999). On Equitable Resource Allocation Problems: A Lexicographic Minimax Approach, Operations Research Vol. 47, No. 3.

Neaga, E.I., Henshaw, M., Yue, Y. (2009). The Influence of the Concept of Capability-based Management on the Development of the Systems Engineering Discipline, Proceedings of the 7th Annual Conference on Systems Engineering Research, 20th - 23rd April 2009, Loughborough University, UK.

Ogryczak, W., Luss, H., Pióro, M., Nace, D., Tomaszewski, A. (2014). Review Article Fair Optimization and Networks: A Survey, Journal of Applied Mathematics, Hindawi Publishing Corporation, 2014, pp.1-26.

Sage, A. P., Cuppan, C. D. (2001). On the Systems Engineering and Management of Systems of Systems and Federations of Systems, Information Knowledge Systems Management 2 (2001) 325–345.

Smith, C. J., Oosthuizen, R. (2011). Applying Systems Engineering Principles Towards Developing Defence Capabilities, Presented at the The Industrial, Systems and Engineering Management conference, Stellenbosch ISEM 2011 Proceedings, South Africa, p. 103.

Smith, C. J., Oosthuizen, R., Harris, H., Venter, J. P.,Combrink, C. and Roodt, J. H. S. (2012). System of Systems Engineering – The Link Between Operational Needs and System Requirements, South African Journal of Industrial Engineering, July 2012, Vol 23 (2): pp 47-60.

Suojanen M., Kuikka, V., Nikkarila, J-P., Nurmi, J. (2015). An Example of Scenario-based Evaluation of Military Capability Areas – An Impact Assessment of Alternative Systems on Operations, IEEE International Systems Conference.

Webb, M. (2008). Capabilities-Based Engineering Analysis (CBEA), Unifying Themes in Complex Systems, pp. 540-547. Springer Berlin Heidelberg.

IV

A Combat Equation Derived from Stochastic Modeling of Attrition Data

Vesa Kuikka

Military Operations Research, 20(3), 2015, pp. 49-69

A Combat Equation Derived from Stochastic Modeling of Attrition Data

Vesa Kuikka

Finnish Defense Research Agency, vesa.kuikka@mil.fi

ABSTRACT

A macroscopic combat equation is derived with stochastic methods. Variables of the equation are the force sizes of the attacker and the defender at the beginning and at the end of the battle. The calculation is based on a combination of force strengths and breakpoints are implied by a win/lose decision methodology. All model parameters of the combat equation have a real world interpretation. Attrition processes are modeled as geometric Brownian motion and stopping rules of battles are modeled as decision boundaries. The theoretical combat equation is compared with historical combat attrition data from an ensemble of different battles and for the evolution of three single battles. The combat equation, with appropriate values of model parameters, agrees with the attrition data. The comparison is accomplished by calculating the probability of victory from the theory and from the empirical data.

INTRODUCTION

In this paper, we discuss macroscopic models that describe battles on a high level. Typical variables of the model are force sizes of an attacker and defender at the beginning and at the end of a battle. Direct effects of weaponry, terrain, or human factors are usually not modeled in macroscopic models. Force sizes govern the situation and they also reflect other variables of the battle. In this paper, we derive combat equations with the help of stochastic modeling. The force sizes of fighting troops are modeled as stochastic processes, for example, Brownian or geometric Brownian motion. Itô's formalism (Karlin and Taylor, 1975) is applied in the treatment of stochastic processes. The model proposed in this paper is applied to empirical data from a large ensemble of different battles, and for the evolution of three historical single battles. The calculation is based on force strengths and win/lose decisions of two opposing sides, and the results are expressed as probabilities of victory.

In the model, the battle ends when the force size of the losing side declines below the lower decision boundary. At this point the losing side surrenders, or the battle ends in some other way. At the same time, the force size of the winning side is assumed to reach the upper boundary value. The decision boundaries are also called stopping boundaries. The calculations can be performed completely with the losing side information and the lower decision boundary. In the model, the winning side decision boundary is needed because of symmetry requirements between two opposing forces. The losing side and the winning side may exchange the roles of losing and winning during the combat. In the course of combat, we could call the decision boundaries upper and lower boundaries because, in fact, the winner is not known before the end of the combat. In addition, the outcome of a battle may be even. These cases are also covered by the proposed model.

In this paper, a simple model for decision boundaries is presented with linear time dependence. More detailed models for the decision boundaries may depend on the environment: casualties, terrain, leadership, troop morale, and so on. Human factors affect the decision process, which makes the modeling more difficult. No empirical data are available, but it may be possible to get expert opinions about the decision parameters and values. Another way for looking at the question is to construct a model for the decision boundaries and to try to make conclusions about the model parameters consistent with the available empirical attrition data.

One successful macroscopic mathematical expression describing the historical data is the Helmbold relationship (Helmbold, 1989). Dean S. Hartley III has fit the parameters of the Helmbold relationship from empirical data of about 800 battles (Hartley, 2001). He has shown that the empirical model holds approximately in different kinds of battles. The model applies to ancient, and in some degree, to modern battles. The author concluded that there may be some fundamental principles behind the Helmbold relationship. There is no theory or modeling explaining the formula.

The combat equation presented in this paper is compared with the empirical Helmbold relationship and the deterministic Lanchester equations (Lanchester, 1914). The Lanchester equations are in common usage considered to be limiting cases of different representations of attrition processes. For example, in the Lanchester

APPLICATIONS AREAS: Modeling, Warfare Performance

OR METHODS: Stochastic Processes, Decision Analysis

square model, all shooters are capable of concentrating their fire on all targets. The new combat equation is more general, because it has more parameters, which makes the model suitable for different combat situations and decision rules, not limited to basic attrition processes described by Lanchester-type equations. On the other hand, in the application part of this paper we demonstrate that analysis of combats can be performed by varying only one or two parameters of the proposed model. Since all the parameters have a real-world interpretation, default values are provided. One of the main differences is that the Helmbold relationship is for the average reference line that separates the empirical data points into two sets. The new formula is for the individual data points. Because the model in this paper is on detailed level, direct comparison of the two models is not possible. Both models predict the winner of the combat in about 79 percent of the cases (Hartley, 2001).

A novelty of our model is in predicting the probability value of a victory. Technically, to calculate the probability of a victory requires calibration of decision boundaries on the bases of information on the combat under investigation or similar other combats. Expert opinions or simulation models can be used in calibration. Secondly, the model can be used as a tool to analyze the effects of weapons and tactical changes of a combat as a function of time. Therefore, greater operational insights and better predictions of battle outcome can be achieved. Attrition data and stopping rules of a combat are used as input for the modeling. We will shortly analyze three historical battles of Kursk, Ardennes, and Inchon-Seul.

In reality, the decision boundary for the losing side to surrender varies during the battle. The simplest model is a linear time-dependent deterministic function. When no reinforcement occurs, the force sizes of the two sides are descending, and this implies descending stopping boundaries. Usually, the two sides don't immediately use all the available force against the enemy. In the model, the effective force sizes may be smaller than the documented force sizes of the historical combat data. If effective force sizes are used the stopping boundaries may be constant, or even ascending, or the slopes of the descending lines may be smaller. The model presented in this paper allows both types of data, real, or effective force size data. Also aggregated data of manpower and weaponry can be used (Bracken, 1995; Turkes, 2000).

Stochastic models don't usually give closed-form formulas for probabilities or force sizes. The method in this paper uses probability theory and the results are presented in closed-form equations. No numeric calculations or no simulation are needed in the analysis. Lanchester equations have stochastic equivalents usually based on discrete time Markov chains (Morse and Kimball, 1951). In practical combat situations, numerical computations of stochastic Lanchester equations are needed because no simple closed-form formula exists.

The combat equation is a function of the initial force sizes, attrition rates, variances, and decision boundaries of two opposing forces. If the attacker or the defender has an extra advantage in the battle, which is not included in other features of the model, a parameter adjusting the decision boundaries is available. This parameter is useful in asymmetric battles.

The combat model of this paper consists of two parts. Firstly, the attrition process is modeled as a stochastic process. Several alternatives exist: Brownian motion, geometric Brownian process, Poisson jump process, and so on (Karlin and Taylor, 1975). Secondly, a model for the decision boundaries is needed. The boundaries can be modeled as deterministic or stochastic functions. In this paper, we use geometric Brownian motion as the model for an attrition process and a linear function of time as the model for decision boundaries. Geometric Brownian motion is one of the few stochastic processes that can be solved in closed-form formulas. Linear decision boundary is a natural choice because the empirical data (Hartley, 2001) do not enable studying more general time dependent functions.

In this paper, the comparison is accomplished by calculating the theoretical probability for the attacker to win the battle from the theory and from the empirical data. Owing to the combat equation, the results are the same if we calculate the probabilities from the defender point of view, even though the parameters for the attacker and the defender may be different.

In the next section, we present a short review of related work of closed-form macroscopic combat models. In the following sections, the stochastic combat model is derived with geometric Brownian motion describing the attrition processes. Also, some justifications for the geometric Brownian motion are provided. Next, a simple model for decision boundaries is presented. Empirical attrition data from about 600 battles is analyzed with the combat model and the model for the decision boundaries. Lastly, data for the evolution of three historical battles are analyzed with the model. In conclusion, a short summary of the proposed model is given.

RELATED WORK IN COMBAT MODELING

A short review of related work and macroscopic combat modeling is presented in this section. Hartley (2001) gives an introduction to the subject and provides more references. The empirical attrition data, used in this paper, is also from Hartley (2001).

Most of the macroscopic combat models are deterministic. The most famous are the Lanchester equations (Lanchester, 1914). Lanchester's equations have been extended in various ways and empirical studies have been published (Chen and Chu, 2001; Fricker Jr, 1998; Hartley, 2001; Jaiswal and Nagabhushana, 1995; Lucas and Dinges, 2004; Salim and Hamid, 2001; Speight, 2001; and Speight 2002). Different forms of the deterministic Lanchester equations give quadratic, linear, and logarithmic descriptions for the force sizes. Lanchester equations have been extensively studied and a goal has been to investigate which of the Lanchester descriptions gives the time-dependent variation of the fighting force size. Historical results of battles do not appear to lead to an unambiguous choice among competing Lanchester equations. Much of the critical effort of the last century has been improving the theory and practice of heterogeneous representation and aggregation, and the representation of the different attrition processes. In spite of its extensive use, the Lanchester formulation suffers from both military (Dupuy, 1985) and mathematical (Ancker and Gafarian, 1992) inconsistencies. The main critique in (Ancker and Gafarian, 1992) is that specific stochastic Lanchester models do not converge to deterministic Lanchester equivalents. The mean outcome of combats is not even defined in relation to a deterministic Lanchester model. In this respect, stochastic models, e.g., stochastic Lanchester models, are better suited with regard to combat modeling. Bonder and Farrell (1970) did the pioneering work on attrition coefficients in heterogeneous target environments. In addition to general inconsistencies of deterministic Lanchester equations, mentioned above, a number of shortcomings exist in Bonner-Farrell methodology, for example, inconsistent treatment for parallel acquisition of targets (Taylor and Neta, 2001). A comprehensive review of combat attrition modeling is presented in Fowler (1995).

Salvo models are types of simple micro tactical models, often based on win/lose ideas, and they exist in both deterministic and stochastic (Hughes, 1995; Armstrong, 2005) forms. The model proposed in this paper is mainly intended for larger battles where statistics, e.g., variances, of attrition processes can be calculated.

Robert L. Helmbold (1989) and Dean S. Hartley (Hartley and Helmbold, 1995) have studied historical battles and they arrived at a result called the Helmbold relationship in Equation (1). The formula gives a relationship between the attacker and the defender force sizes at the beginning and at the end of the battle. The Helmbold relationship does not depend explicitly on any Lanchester attrition laws.

$$\ln\left(\frac{x_0^2 - x^2}{y_0^2 - y^2}\right) = \alpha \ln\left(\frac{x_0}{y_0}\right) + \beta, \qquad (1)$$

where $x = x(t)$ and $y = y(t)$ are force sizes of the opposing forces at time t and $x_0 = x(0)$ and $y_0 = y(0)$. In Equation (1), α and β are constants. The logarithm of the Helmbold ratio (left side of Equation (1)) as a function of the logarithm of the initial force ratio separates the empirical data into two parts (Hartley, 2001). Below the ascending line the attacker is more probable to win and above the line the defender wins more probably. Partly, this can be explained by the

different numbers of casualties of attackers and defenders. This cannot explain the power law of the relationship or the values of the parameters. The data points don't follow the equation exactly because it is an average description of the situation.

One of the earliest stochastic models has been published in (Morse and Kimball, 1951). In the model, the probabilities of the time-dependent casualties have been calculated for the attacker and the defender. Stochastic war equations have been studied in (Hausken and Moxnes, 2002, 2005) and stochastic variations of Lanchester models have been discussed in Kingman (2002). There are also agent-based (Hill et al., 2003), fractal-based (Lauren, 2002), and cellular automata based (Moffat et al., 2006) approaches.

The basic Lanchester laws don't have a rule to end the battle, except when the force size goes down to zero on one side of the battle. In Jaiswal and Nagabhushana (1995) an ending rule has been used, where absolute or proportional force sizes determine the boundary, and expressions for the force sizes at the end of the battle have been derived. The effect of information in military operations has been studied in Perry (2003).

Hartley's (1995) study shows that the Helmbold relationship may be explained by constraints on the force sizes. This is related to the model in this paper.

Combat models can be static or dynamic, i.e., time independent or time dependent. The first static models were based on numbers of troops and weapons and their appropriate weighting factors. Comparing the weighted sum of the two opposing sides predicts the winner of a battle or a war. T. N. Dupuy (1985) developed the static Quantified Judgment Model (QJM), which takes into account, in addition to weaponry and force sizes, human factors such as leadership, troop morale, and operative factors like terrain and weather. Parameters of the model have been calculated with the help of theoretical and empirical results. In QJM, a power potential value for both sides is calculated, and the ratio of the two values predicts the victor. For example, if A's power potential exceeds D's power potential, the model predicts that A is the victor:

$$\frac{P(A)}{P(D)} > 1.$$

The most common and the most studied family of combat models are the Lanchester equations (Lanchester, 1914). The Lanchester differential equations describe the force sizes in the battle as a function of time. Different battles have been described with quadratic, linear, and logarithmic Lanchester differential equations. The equations model the attrition of the two forces. In the quadratic model the derivative of force sizes on both sides of the battle are proportional to the force sizes of the opposing side:

$$\frac{dx}{dt} = -D_{qy}y \text{ and } \frac{dy}{dt} = -D_{qx}x. \tag{2}$$

Equation (3) gives the solution to the differential equations where D_{qx} and D_{qy} are constants.

$$\frac{D_{qx}}{D_{qy}} = \frac{(x_0^2 - x^2)}{(y_0^2 - y^2)}, \tag{3}$$

where $x = x(t)$ and $y = y(t)$ are forces sizes of the opposing forces at time t and $x_0 = x(0)$ and $y_0 = y(0)$. Equation (3) is quadratic with respect to the force sizes. Equations (2)–(3) hold also as a function of time during the engagement. The quadratic Lanchester equation depicts a modern combat with firearms engaging each other directly with aimed shooting from a distance, which can attack multiple targets and can retrieve fire from multiple directions. For the linear Lanchester model the corresponding equations are:

$$\frac{dx}{dt} = -D_{lx}xy \text{ and } \frac{dy}{dt} = -D_{ly}xy,$$

$$\frac{D_{lx}}{D_{ly}} = \frac{(x_0 - x)}{(y_0 - y)}.$$

The linear Lanchester equation depicts unaimed fire into an enemy-occupied area. The linear law also applies to target acquisition situations. And for the logarithmic law the equations are:

$$\frac{dx}{dt} = -D_{gx}x \text{ and } \frac{dy}{dt} = -D_{gy}y,$$

$$\frac{D_{gx}}{D_{gy}} = \frac{(\ln x_0 - \ln x)}{(\ln y_0 - \ln y)}.$$

The logarithmic law is typically used to describe the number of casualties of nonfighting participants, such as medical personnel and headquarters staff. The analysis of the Helmbold relationship, Equation (1), yielded a value for α of 1.3 indicating a Lanchester law between the linear and the logarithmic laws, but closer to the linear law (Hartley, 2001).

Willard has examined about 1,000 battles from years 1618–1905 and has suggested the following empirical result (Willard, 1962):

$$\ln\left(\frac{x_0 - x}{y_0 - y}\right) = \ln E + \gamma \ln\left(\frac{x_0}{y_0}\right),$$

where E and γ are constants. We can compare this with the Helmbold relationship, which is also based on empirical studies (Hartley and Helmbold, 1995; Hartley, 2001; Hartley, 1995; Helmbold, 1989).

Dean S. Harley III showed that the Helmbold relationship is in good agreement with the observed data. He used Land Warfare Database (LWDB) that has the combat data of about 600 battles and, in addition, he used the data of about 200 battles from other sources (Hartley, 2001).

Dean S. Harley III has developed his own model called Oak Ridge Spreadsheet Combat Model (ORSBM) (Hartley, 2001). The model includes force sizes, weapons, human factors, operational variables, and environmental variables as independent variables. Most important predicted variables are the force sizes at the end of the battle and the measure of success.

The new combat equation of this study is presented in Equation (38).

STOCHASTIC INTRODUCTION

In this section we give a short introduction to stochastic processes following (Karlin and Taylor, 1975). The results will be used in the following sections. Later on, the force sizes of the opposing forces are modeled as Brownian or geometric Brownian motions.

Consider a Brownian motion B_t with constant drift $\mu \in R$ and constant standard deviation $\sigma \in R^+$ where R is the set of real numbers and R^+ the set of positive real numbers. The stochastic process B_t follows the stochastic differential equation:

$$dB_t = \mu dt + \sigma dW_t, \tag{4}$$

where t is time and W_t is a standard Brownian motion. Standard Brownian motion at time t is normally distributed with mean zero and variance t: $W_t \sim N(0, t)$. Equation (4) can be solved by integration:

$$B_t = B_0 + \mu t + \sigma W_t, \tag{5}$$

where B_0 is a constant. The geometric Brownian motion X_t can be constructed from the process B_t:

$$dX_t = X_t dB_t = X_t \mu dt + \sigma X_t dW_t. \tag{6}$$

Stochastic differential Equation (6) can be solved with the help of Itô formula (Karlin and Taylor, 1975) giving

$$X_t = e^{t\left(\mu - \frac{1}{2}\sigma^2\right) + \sigma W_t}.$$

Later, we call this the log-normal distribution.

Next, we return to the Brownian motion B_t in Equation (4). We examine the stochastic process B_t in the interval $[d, u]$. The stopping time T is the time when B_t reaches the upper or the lower boundary. We consider boundaries d and u, which are constant or linear functions of time. For simplicity, the same slopes for the upper and the lower boundaries are assumed. In a more general case, the boundaries can be stochastic functions of time.

The proof of Theorem 1 with constant boundaries is well known (Karlin and Taylor, 1975). Here, we present the proof for linear time dependent boundaries.

Theorem 1. *Let B_t be a Brownian motion (5) with parameters $\mu \in R$ and $\sigma \in R^+$, where $B_0 = b$. Let $D < B_0 = b < U$ and*

$$u(t) = U - Mt \text{ and } d(t) = D - Mt, \tag{7}$$

where U, D, and M are constants and

$$T \equiv T_{du} = \inf\{t > 0 | B_t = d(t) \vee B_t = u(t)\}. \tag{8}$$

is the stopping time for the Brownian motion B_t. The probability for the process to reach first the upper boundary u when starting from b is

$$P(B_T = u(t) | B_0 = b) = \frac{e^{mb} - e^{md}}{e^{mu} - e^{md}} = \frac{e^{mb} - e^{m(D+MT)}}{e^{m(U+MT)} - e^{m(D+MT)}}, \tag{9}$$

where

$$m = \frac{-2(\mu + M)}{\sigma^2}. \tag{10}$$

Proof. First, we consider constant boundaries D and U $(M = 0)$. A new stochastic process V_t is defined (Karlin and Taylor, 1975):

$$V_t = e^{mB_t - \left(m\mu + \frac{1}{2}m^2\sigma^2\right)t}.$$

V_t is a martingale for all m, i.e., $E(V_t \mid V_s) = V_s$ holds for the conditional expected values. We choose the value of m that makes the second term zero: $m_0 = -2\mu/\sigma^2$. By using the martingale property and the optional stopping theorem (Karlin and Taylor, 1975) we have for the probabilities

$$e^{m_0 b} = E(V(T_{du})) = P(B_{T_{du}} = U)e^{m_0 U} + P(B_{T_{du}} = D)e^{m_0 D}.$$

Because $P(B_{T_{du}} = U) = 1 - P(B_{T_{du}} = D)$ is the probability for the process to reach first the boundary u when starting from b we have

$$P(B_{T_{du}} = U | B_0 = b) = \frac{e^{\frac{-2\mu}{\sigma^2}b} - e^{\frac{-2\mu}{\sigma^2}D}}{e^{\frac{-2\mu}{\sigma^2}U} - e^{\frac{-2\mu}{\sigma^2}D}}. \tag{11}$$

When the stopping boundaries are deterministic linear functions of time the parameter m_0 transforms into

$$m = \frac{-2(\mu + M)}{\sigma^2},$$

and we get Equation (9) from Equation (11) in the case of time dependent boundaries. ■

If $m = 0$ in Equation (9) l'Hôpital's rule gives

$$P(B_{T_{du}} = u(T) | B_0 = b) = \frac{b - D}{U - D}. \tag{12}$$

In practice, the martingale property used in the proof requires that no information about the future movement of the process V_t is known after time s and the relationship $E(V_t \mid V_s) = V_s$ follows from the properties of Brownian motion. In combat situations, if actions of either side are known in advance, the martingale property is not valid. In this paper, we assume that the martingale property is a sufficient approximation for the force sizes. The martingale property suggests a way of practical application of the theory. The closed-form combat equation provides a tool for forecasting possible future results, as well as for analyzing past results.

In the transformation from constant to linear time dependent boundaries two things happen: The slope of the Brownian motion is increased by M (M can be negative) and the boundary values are reversed to the initial constant values of the boundaries: $U = u(t) + Mt$ and $D = d(t) + Mt$. The second equality in (9) follows from this, when M is not equal to zero.

DERIVATION OF THE COMBAT EQUATION

In this section a combat equation is derived based on stochastic theory. We perform this in two phases. In the first phase the attrition processes are modeled as Brownian motions and in the second phase as geometric Brownian motions.

The two opposing sides are called the attacker and the defender. We follow the convention in Hartley (2001) because the attrition data is given in this order, first for the attacker, and second for the defender.

The force size of the attacker is modeled by the Brownian motion $B_{A,t}$, and the force size of the defender by the Brownian motion $B_{D,t}$.

Theorem 2. *Assume that the force size of the attacker (A) and the defender (D) are Brownian motions of Equation (4) with constant parameters $\mu_A \in R$, $\sigma_A \in R^+$, and $\mu_D \in R$, $\sigma_D \in R^+$, respectively. We assume*

a) *the probability for A to win the battle equals the probability for D to lose the battle*
b) *the upper and lower stopping boundaries for the attacker are linear in time $a_u - M_A t$ and $a_d - M_A t$, where a_u is the initial upper decision boundary, a_d is the initial lower decision boundary and M_A is the constant slope. The corresponding boundaries for the defender are $d_d - M_D t$ and $d_d - M_D t$.*
c) *the stopping time for the attacker and the defender is*

$$\begin{aligned} T &= \{\min(t) | B_{A,t} = a_u - M_A t \vee B_{A,t} = a_d - M_A t\} \\ &= \{\min(t) | B_{D,t} = d_u - M_D t \vee B_{D,t} = d_d - M_D t\}. \end{aligned} \tag{13}$$

The following combat equation follows:

$$\frac{e^{m_A a_0} - e^{m_A(B_{A,T_u} + M_A T_u)}}{e^{m_D d_0} - e^{m_D(B_{D,T_d} + M_D T_d)}} = \frac{e^{m_A a_0}}{e^{m_D d_0}} \left(\frac{1 - \dfrac{e^{m_A(B_{A,T_d} + M_A T_d)}}{e^{m_A a_0}}}{1 - \dfrac{e^{m_D(B_{D,T_u} + M_A T_u)}}{e^{m_D d_0}}} \right), \tag{14}$$

where a_0 and d_0 are the initial attacker and defender force sizes and $B_{A,T}$ and $B_{D,T}$ are the values of the attacker and defender force sizes at stopping time T.

The exponents of Equation (10) for the attacker and the defender are

$$m_A = \frac{-2(\mu_A + M_A)}{\sigma_A^2}, m_D = \frac{-2(\mu_D + M_D)}{\sigma_D^2}. \tag{15}$$

Proof. When the process stops at the lower boundary, we denote the stopping time of Equation (13) T_d and the value of the process B_{A,T_d} for the attacker and B_{D,T_d} for the defender. When the process stops at the upper boundary we denote the stopping time T_u and the value of the process B_{A,T_u} for the attacker and B_{D,T_u} for the defender. Because of the assumption c) in Theorem 2, $T = T_u = T_d$.

When the probability for A to win the battle equals the probability for D to lose the battle, then

$$\begin{aligned} &P(B_{A,T_u} = a_u - M_A T_u | B_{A,0} = a_0) \\ &\quad = P(B_{D,T_d} = d_d - M_D T_d | B_{D,0} = d_0). \end{aligned} \tag{16}$$

We use the following shorthand notations

$$\begin{aligned} A_0 = e^{m_A a_0}, \quad A_d = e^{m_A(B_{A,T_d} + M_A T_d)}, \\ A_u = e^{m_A(B_{A,Tu} + M_A T_u)}, \\ D_0 = e^{m_D d_0}, \quad D_d = e^{m_D(B_{D,T_d} + M_D T_d)}, \\ D_u = e^{m_D(B_{D,Tu} + M_D T_u)}. \end{aligned}$$

Equation (16) can be transformed with Equation (9) into

$$\frac{A_0 - A_d}{A_u - A_d} = \frac{D_u - D_0}{D_u - D_d}. \tag{17}$$

Further, this gives

$$\frac{A_0 - A_u}{D_0 - D_d} = \frac{A_0 - A_d}{D_0 - D_u} = \frac{A_0}{D_0} \left(\frac{1 - \dfrac{A_d}{A_0}}{1 - \dfrac{D_u}{D_0}} \right). \tag{18}$$

The unknown values A_d and D_u that have not been observed are in the parenthesis. Equation (14) follows from Equation (18). ■

Equation (14) brings together attacker and defender sides' variables of the model. In literature (Perry, 2011), difference has been made between models of combat and models of combat attrition. For example, Lanchester's equations are a model for attrition and they cannot be expected to capture other effects of engaged forces. Equation (14) describes both attrition and decision rules for both sides of the combat. Itô's formalism has been studied in (Perry, 2011) in relation with the Helmbold relationship but no decision methodology has been used. In our paper, a combat equation is derived from stochastic attrition processes with Itô's formalism and decision boundaries. In this respect, our approach is more fundamental. One conclusion in (Perry, 2011) is that an additional term (e.g., in the Helmbold relationship) appears as a result of Itô's formalism. This is a substantial difference between stochastic and deterministic modeling of processes and, as a consequence, different results are expected from stochastic and deterministic methods.

This combat equation holds if the attacker and the defender force sizes are modeled by Brownian motions. The values of A_u and D_d are known and these quantities are moved to the left side of the equation.

Next, we study another alternative. Assume that the force sizes of the attacker and the defender have log-normal distributions (the process is a geometric Brownian motion). By definition, the stochastic variables X and Y have log-normal distributions when

$$B_{A,t} = \ln(X_t) \text{ and } B_{D,t} = \ln(Y_t), \tag{19}$$

where $B_{A,t}$ and $B_{D,t}$ are normally distributed.

The log-normal distribution is multiplicative, which means that on a future state X_t the process value depends on a previous value

X_{t-1} according to the following linear transformation (Mitzenmacher, 2004):

$$X_t = F_t X_{t-1}. \tag{20}$$

If the factors F_t are independently log-normally distributed, consequently, X_t are log-normally distributed, because the product of log-normal stochastic variables is log-normally distributed. The central limit theorem (Karlin and Taylor, 1975) asserts that X_t is log-normally distributed in a more general situation: especially if F_t are independent identically distributed stochastic variables, the process X_t approaches asymptotically the log-normal distribution (Mitzenmacher, 2004).

Equations (6) and (20) are common in modeling phenomena like stock values. Stochastic changes in these quantities are proportional to the values of the quantity itself. This can be justified in combat modeling also. Deterministic Lanchester logarithmic law obeys a similar rule. We could model attrition processes with stochastic processes comparable with linear and quadratic Lanchester laws. This is a possible future study, and closed-form formulas may not be attained.

Denote the initial values of the log-normally distributed stochastic variables X and Y by x_0 and y_0, and the final known force sizes by x_u and y_d and the final unknown force sizes by x_d and y_u. The final values are at stopping time T.

Theorem 3. *Assume that the force size of the attacker (A) and the defender (D) are geometric Brownian motions (6) with constant parameters $\mu_A \in R$, $\sigma_A \in R^+$, and $\mu_D \in R$, $\sigma_D \in R^+$, respectively. Make the assumptions a–c) of Theorem 2. The following combat equation follows*

$$\frac{x_0^{m_A} - e^{m_A M_A T} x_u^{m_A}}{y_0^{m_D} - e^{m_D M_D T} y_d^{m_D}} = \frac{x_0^{m_A}}{y_0^{m_D}} \left(\frac{1 - \dfrac{e^{m_A M_A T} x_d^{m_A}}{x_0^{m_A}}}{1 - \dfrac{e^{m_D M_D T} y_u^{m_D}}{y_0^{m_D}}} \right). \tag{21}$$

Proof. Equation (21) follows from Equations (14) and (19).■

Equation (21) holds if the attacker and the defender force sizes are modeled by log-normal distributions. These can be compared with Equations (14) where the force sizes have been modeled by Brownian motions. In fact, it is possible that in an asymmetric battle the attacker and the defender force sizes obey different stochastic laws, and a mixed form of the combat equations follows.

In Figure 1, an example is given for the time dependencies of the attacker and the defender force sizes. The battle ends when the defender force size intersects the lower linear boundary. At the same time, the attacker force size is assumed to intersect the upper boundary. If the defender is victorious the situation is opposite. In reality, the boundaries don't chance independently because both sides evaluate their positions and make changes in their tactics. The linear boundary can only be an approximate model.

One could argue that only one of the boundaries exists in reality, namely the losing side boundary, and the other boundary is imaginary. Later, we will investigate time dependent combat datasets from three historical battles. We use either the attacker or the defender view in the calculations. We show that both ways give the same results with appropriate parameter values for the decision boundaries. We can always take the losing side view in the calculations to make sure that we are working with the real world process. As said, both ways still give the same results.

Time-dependent boundaries can be a consequence of changing decision rules or reinforcement, or both. Calculations can be

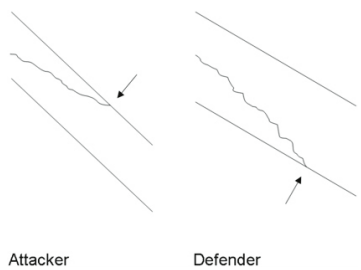

Figure 1. On the left an example for the attacker force size and on the right an example for the defender force size are shown as a function of time. The arrows indicate where the attrition processes hit the decision lines.

carried through with effective, smaller force sizes with reinforcement or, equivalently, with total force sizes. Both approaches should give the same results. In addition, different degrees of casualties, killed, wounded, etc., and also weaponry, can be taken into account with weighting factors in the attrition process. The stopping boundaries are different in all these cases. Typically the decision boundaries are descending, but in reinforcement situations, the decision boundaries may be constant or ascending.

A deficiency of the proposed model is that the attrition process and the reinforcement process may not be symmetric. As a result, the geometric Brownian motion is only an approximation for a more realistic combined process describing both attrition and reinforcement. However, using declining decision boundaries and appropriate parameter values can partly compensate for the problem.

A MODEL FOR THE DECISION BOUNDARIES

In Equation (21) the decision boundaries M_A and M_D are not specified and our next goal is to build a model for the boundaries. We can use other processes of the model. In fact, we have only two candidates, the attrition processes of the attacker and the defender. Our first idea is to model $M_A = -\mu_D$ and $M_D = -\mu_A$. The attrition rates μ_A and μ_D are negative for decreasing force sizes. The minus sign is a convention of previous. This choice has favorable properties. The situation is symmetric for the attacker and the defender. Secondly, the conditions $M_A > -\mu_A$ and $M_D > -\mu_D$ hold if $-\mu_D < -\mu_A$ (Figure 1). To have a more general model we assume that

$$M_A = -(1-\alpha)\mu_D - \alpha\mu_A, \tag{22}$$

$$M_D = -(1-\beta)\mu_A - \beta\mu_D, \tag{23}$$

where α and β are constants. Intuitively we expect that α and β are close to one. Different parameters give a possibility to model an advantage for the attacker, for example. In previous research, the basic models are often symmetric between the attacker and the defender. On average, the attacker has a minor advantage in empirical data, but sometimes better weaponry, leadership, or other conditions can lead to a situation where combat models need a method to model the advantage for the defender side. If we assume geometric Brownian motion for the attrition process, the attrition rates are:

$$\mu_A = \ln\frac{x_u}{x_0}, \tag{24}$$

$$\mu_D = \ln\frac{y_d}{y_0}, \tag{25}$$

where x_u and y_d are the values of force sizes at the end of the combat for the attacker and the defender. The initial values x_0 and y_0 are the corresponding values at time 0. In the historical combat data, the force sizes are known at the beginning and at the end of the combat. Usually time dependency of the process is not known and most of the data are for one day duration. Thus the variances of the attrition processes cannot be calculated from the data and we have to take variation of a processes as a free parameter. However, there are some time-dependent data available and we can calculate the variances for those combats. It is important to check that the parameter values are of the same order of magnitude compared to the parameter values consistent with rest of the empirical data.

We take the duration of the battle as our time unit. This means that the attrition rates and the variances are calculated for the duration of the combat. In the case of geometric Brownian motion, the variances are calculated from the logarithmic data and are commonly called volatilities. If time-dependent data is available, daily volatilities can be calculated and, if needed, the values can be transformed to another time unit.

From Equations (22–23) we get

$$M_A = -\alpha\ln(S) + \ln\left(\frac{y_0}{y_d}\right), \tag{26}$$

$$M_D = \beta\ln(S) + \ln\left(\frac{x_0}{x_u}\right), \tag{27}$$

where

$$S = \frac{x_u y_0}{x_0 y_d}. \tag{28}$$

From Equations (15), (24–25), and (26–27) we get for the exponents:

$$m_A = \frac{-2}{\sigma_A^2}(1-\alpha)\ln(S), \tag{29}$$

$$m_D = \frac{-2}{\sigma_D^2}(\beta-1)\ln(S). \tag{30}$$

If the opposing forces have the same attrition variances and the same parameters α and β, we have $m_D = -m_A$, and if $\ln(S)$ is positive and $\alpha < 1$, m_D is positive and m_A is negative.

Now we consider the situation in Figure 2. The initial value of the lower boundary is

$$x_{d0} = x_0 - p_1 L = p_1(e^{M_A} x_u - x_0), \tag{31}$$

and the final value of the lower boundary is

$$x_d = e^{-M_A} x_{d0} = e^{-M_A} x_0 - p_1(x_u - e^{-M_A} x_0) = e^{-M_A} x_0(1+p_1) - p_1 x_u = (S^{\alpha-1}(1+p_1) - p_1)x_u, \tag{32}$$

where we used the relationship

$$e^{-M_A} x_0 = S^{\alpha}\frac{y_d x_0}{y_0 x_u} x_u = S^{\alpha-1} x_u.$$

In the equations the value of the upper boundary is the same as the final value of the attacker force size if the attacker wins. The next equation

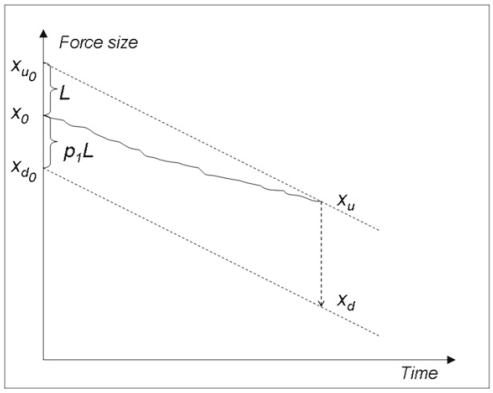

Figure 2. Variables of Equation (31).

$$x_d = A x_u \tag{33}$$

gives the final value of the lower boundary where

$$A = S^{\alpha-1}(1+p_A) - p_A \tag{34}$$

is a function of parameter p_A. The parameter determines the initial position of the lower boundary and describes the initial advantage or disadvantage of the attacker. Values greater than one stand for advantage and values less than one for disadvantage.

A very concrete decision boundary exists if a fighting force has a doctrine level instruction to surrender after specific conditions of the combat.

When calculating the probability for the attacker (or the defender) to win, we have to elaborate Equation (17) in more detail. Equation (17) gives the probability for the attacker to win if $S > 1$. This is the situation depicted in Figure 1. But if $S < 1$, the situation in Figure 1 is reversed.

Theorem 4. *If $S > 1$, the probability for the attacker to win is*

$$P_A = \frac{A_0 - A_d}{A_u - A_d} = \frac{(x_0^{m_A} - e^{m_A M_A} x_u^{m_A})}{(e^{m_A M_A} x_u^{m_A} - e^{m_A M_A} x_d^{m_A})}. \tag{35}$$

If $S < 1$, the probability for the attacker to win is

$$P_A = \frac{(x_0^{-m_A} - e^{-m_A M_A} x_u^{-m_A})}{(e^{-m_A M_A} x_u^{-m_A} - e^{-m_A M_A} x_d^{-m_A})}. \tag{36}$$

Proof. When $S < 1$, Equation (35) follows from Equations (16)–(17) and the situation in Figure 1. If $S < 1$, the probability for the attacker to win is still the probability to reach the upper boundary. Because the order of the lower and the upper boundaries has been exchanged, the probability for the attacker to win is obtained by reversing the sign in m_A. ■

Equations (35)–(36) can be expressed shortly as

$$P_A = \frac{(x_0^{-|m_A|} - e^{-|m_A| M_A} x_u^{-|m_A|})}{(e^{-|m_A| M_A} x_u^{-|m_A|} - e^{-|m_A| M_A} x_d^{-|m_A|})}. \tag{37}$$

Theorem 4 is given for the attacker, but the corresponding theorem holds for the defender. In the next sections we choose the attacker view. Both the attacker view and the defender view give the same results. The parameters of the

attacker process and the defender process are connected by the combat Equation (21). The expression in Equation (37) is sufficient for comparing the empirical attrition data and the theory with a model for decision boundaries.

Theorem 5. The probability for the attacker to win is

$$P_A = \frac{(S^{\alpha-1})^{-|m_A|} - A^{-|m_A|}}{1 - A^{-|m_A|}}, \tag{38}$$

where S, m_A, and A have been given in Equations (28), (29), and (34), correspondingly.

Proof. Equation (38) follows from Equations (26), (28), and (37). ■

The corresponding expression for the defender to lose from the right side of Equation (17) is

$$P_D = \frac{(S^{1-\beta})^{|m_D|} - D^{|m_D|}}{1 - D^{|m_D|}}, \tag{39}$$

where

$$D = S^{1-\beta}(1 + p_D) - p_D. \tag{40}$$

In theory, these probabilities are equal. This means that the parameters of the attacker and defender sides are related by the equation:

$$P_A = P_D. \tag{41}$$

We call these the attacker view and the defender view of the combat. In data analysis we use the attacker view, but both views give the same results, because the parameters are synchronized according to the Equation (41).

Later we need the value of P_A or P_D when S approaches one. l'Hôpital's rule gives

$$\lim_{s \to 1} P_A = \frac{p_A}{1 + p_A} \tag{42}$$

and

$$\lim_{s \to 1} P_D = \frac{p_D}{1 + p_D}. \tag{43}$$

Derivatives of P_A and P_A are zero, which is a natural result because of the symmetry between the two sides of the combat when S equals one.

When S equals one, from Equations (41)–(43) we get

$$p_A = p_D.$$

This is one example of relationships between the parameters of the attacker and the defender in the combat equation.

COMPARISON WITH THE ATTRITION DATA

The combat data (Hartley III, 2001) has 593 records of historical combats. A typical record of the data is: "A, 4000, 21000, 400, 5000." The first field has the information of the winner of the combat: attacker A or defender D. This information is not used in the theory. The next four data fields are initial force size of the attacker, initial force size of the defender, casualties of the attacker, and casualties of the defender. Only these four data fields have been used in this study.

The combat equations are symmetric for the attacker and the defender. From the decision boundary model of Equations (22)–(23) follows that the probability in Equation (38) is a function of S in Equation (28). The quantity S is a good indicator of the victor of a combat. If $S < 1$, the defender is predicted to win and if $S > 1$, the attacker is predicted to win the combat. Other models, the Helmbold relationship, the Willard relationship or the empirical formula (45) give almost the same number of correct predictions for the victor. For example, with the values above the numerical value of S is:

$$S = \frac{(4000 - 400) \times 21000}{4000 \times (21000 - 5000)} = 1.18$$

The prediction is that the attacker wins the battle. Indeed, this is the case because the first data field has the value A. In the data set of 593 records 354 attacker wins, 198 defender wins and 41 are even.

All the three models, and the model in this paper, predict the same 438 records correctly and 114 incorrectly. In some cases the value of S is very near or exactly 0.5 and some parameters of the model may give 438–441 correct predictions. This cannot be used to calibrate the model parameters, because of the variations of the empirical data and combat situations. The conclusion is that it is easy to give an empirical formula predicting the winner in about 438 of 552 cases. The even 41 cases are not included in this comparison.

At this stage we used only the ratio S, which is the primary quantity of the combat model predicting the victor in this paper. It is a consequence of the decision boundary model of the previous section. Another model for the decision boundary may result in a different result. We tried the following model of Equation (44). No better agreement with the empirical data was achieved, and that is why the results are not presented here. In Equation (44) α and β are constants.

$$M_A = -(1-\alpha)\mu_D - \beta\mu_A. \tag{44}$$

Our model gives the probability of victory. We investigate which parameters of the model agree with the empirical attrition data. In Hartley III (2001) an empirical formula for the probability of attacker victory has been presented:

$$P = 1 - \frac{1}{1+e^{\bar{q}}} \tag{45}$$

where

$$q = -1.75 sign(v)|v|^{0.75}$$

and

$$v = \ln\left(\frac{x_0^2 - x^2}{y_0^2 - y^2}\right) - 2\ln\left(\frac{x_0}{y_0}\right),$$

where x_0 (y_0) and x (y) are the values of force sizes at time 0 and at the end of the combat for the attacker (defender). The formula is not accurate because of the variable data and the different combat situations. The error in the formula is 5 percent when compared with empirical data. We compare the values of P_A in Equation (38) and the values of Equation (45) with the following simple method:

$$\Delta = \frac{1}{593}\left(\sum_{i=1,593} |P_A(i) - P(i)|\right). \tag{46}$$

The parameters of the model and their interpretations have been summarized in Table 1. Parameter values for the attacker and defender may be different.

A wide range of parameter values of the theory agree with the empirical probability formula (45). In Table 2 some representative parameter values have been listed. They predict correctly 427–441 victors of the combat data (column # in the table). Good compliance with the empirical data occurs when the number of correctly predicted victors is 437 or more. The Δ values vary between 0.137 and 0.159. All values of Δ in Table 2 can be regarded as good. No more conclusions can be made because the accuracy of Equation (45) is poor or unknown.

Table 1. List of the parameters of the model.

σ^2	Variance of logarithmic attrition data (volatility).
α	The parameter determines the slope of a decision boundary.
p_A	The advantage parameter of a decision boundary.

The parameters are not independent and it is not meaningful or possible to fit the parameters. Nevertheless, some conclusions can be drawn. In Table 2, the parameters for the attacker view are shown. The defender has the corresponding parameters connected by Equation (41).

The model does not limit the values of variance. Small values correspond to high absolute value of m_A in Equations (29)–(30). In theory, the variance values may vary from zero to infinity. In the second part of Table 2, example values of quantities A and m_A are shown for some typical values of S. Later in this paper, we present how the parameter values α and p_A can be evaluated by using expert opinion. The value of variance σ^2 is determined from the time dependent data of a combat (if available).

Table 2. Typical parameter values of σ^2 and p_A ($\alpha = 0$) of the model with correctly predicted numbers of victors (#) and values of Δ from Equation (46). In the second part of the table values of A and m_A from Equations (34) and (29) for some typical values of S from Equation (28) are shown.

σ^2	p_A	#	Δ	S	A	m_A
0.00025	1	438	0.150	1	1.0	0.0
0.0025	1	438	0.136	1.05	0.905	−39
0.025	1	438	0.149	1.1	0.818	−7.6
0.00025	1.1	441	0.142	1	1.0	0.0
0.0025	1.1	441	0.137	1.05	0.9	−39
0.025	1.1	432	0.147	1.1	0.809	−7.6

Because the data set (Hartley, 2001) has only the initial and final values of force sizes, variances cannot be calculated from the data. However, time-dependent data from three historical combats are available (Armstrong, 2005; Turkes, 2000; Hartley and Helmbold, 1995): Ardennes, Kursk, and Inchon-Seoul combat data. Variances calculated from these three cases are listed in Table 3. The variances are calculated as variances of logarithmic attrition data. In financial applications this quantity is called volatility. The variances are shown for the attacker and for the defender. The variances for the duration of the battle are high compared with the typical values of Table 2. The variances for one day can be easily obtained by dividing the variance values in Table 3 by the duration of the battle. The attrition rates and the volatilities must be calculated with the same time unit.

Next, we have the decision boundary variable α, which may be interpreted as a measure of resilience. It describes how steep the boundary line is. We expect that usually the boundary line is steeper for the winner of the combat which is the idea behind Equations (22)–(23). Low values of α give a better correspondence with the empirical data. Negative values of α may be possible in some situations, but normally α is positive or zero. In general, the model is not very sensitive to changes of α values.

Finally, the parameter p_A describes the initial advantage of the attacker (Figure 2). We expect intuitively that, on the average, p_A is close to one. Only positive values of p_A are allowed in the model. From Table 2 we see that the value of $p_A = 1.1$ gives better results. This indicates that the attacker has a small advantage and this agrees with earlier research (Hartley, 2001).

In Figure 3, theoretical values from Equation (38) are shown with parameter values $\alpha = 0$, $p_A = 1.1$ and three variance values indicated in the figure. In Equations (34) and (38), the advantage for the attacker can be taken into account by giving a value above one for the parameter p_A, for example $p_A = 1.1$. This shifts the theoretical curves in Figure 3 up 1–3 percent when S is between 0.95 and 1.05. As a conclusion, a good compliance with the empirical data is obtained with the value $p_A = 1.1$. Higher values of p_A don't make the fit better. Variance value 0.00025 for values of S between $0.98 < S < 1.02$ and variance value 0.0025 for values of S between $1.02 < S < 1.05$ give a good agreement with the average empirical data. When $S < 0.98$ and $S > 1.05$ even higher values of variance agree better with the data. This can be a consequence of more attrition and more intensive battles. Because the theory holds for individual data points, all the data can be consistent with the theory with appropriate parameter values.

Table 3. Variances of historical attrition data. Variance values are for the duration of the combat. Duration is shown in the last column. Combat power includes manpower and weaponry with weighting factors. T4, T5, and T15 refer to Tables 4 and 5 (in Appendix).

Battle	Attacker variance	Defender variance	Duration in days
Kursk (T15)	0.0029	0.098	15
Ardennes manpower (T4)	0.335	0.096	33
Ardennes combat power (T5)	0.400	0.156	33
Inchon-Seul	0.024	0.160	19

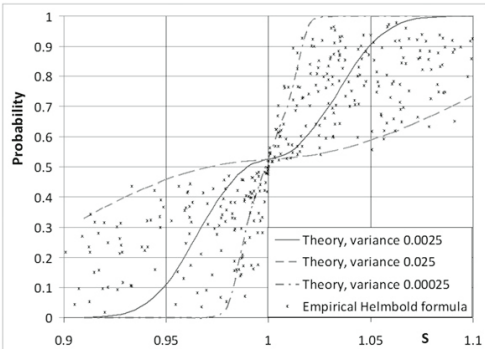

Figure 3. Theoretical probability values from Equation (38) with parameter values $\alpha = 0$ and $p_A = 1.1$ and variance values indicated in the figure. Empirical Helmbold results are calculated from Equation (45) for empirical data points.

AN APPLICATION OF THE MODEL ON THE EVOLUTION OF A BATTLE

In this section a practical application of the theory of previous sections is presented. First, we present the main steps of the analysis. Force and attrition data on both sides of the battle are assumed to be available. The analysis is easily conducted with a spreadsheet application. Evaluating the possible advance position and decision mechanisms requires more analytic skills. The most demanding task is the calibration of the level of probability. From historical battles the outcome is known and the information can be used for calibration. In practice, this is sufficient for the analysis, because the overall structure of the probability curve is not affected. In Table 4 (in Appendix), instructions are provided for adjusting the advance parameter and the decision boundary parameters. In previous sections, variance is a model parameter because variance values are not known for the combat data. In this section we use two-sided observations on battle days, which make it possible to calculate the variances from empirical data.

Before the analysis of historical battles, we examine three example curves with different parameter values. The example data is from the battle of Kursk (Hartley, 1995). Figure 4 shows how the parameter values p_A (and p_D) and α (and β) in general affect the probability curves. Probabilities are calculated for 15 days of the battle from Equation (38). The parameter p_A simply shifts the level of the curve. The curve with $p_A = 1.1$ gives 2–3 percent higher probabilities depending on the value of S (curve b compared to curve a). When S is close to 1, Equation (42) provides the relation between p_A and the probability value. The decision boundary parameter α affects both the level and the shape of the curve as can be seen in Figure 4. When $S < 1$ raising the value of α from 0.1 to 0.4 gives higher probability values (curve c compared to curve b) and when $S > 1$, this gives lower values. During days 1–3 $S > 1$, and curve c is below curve b. Also the shape in vertical direction is flatter with higher values of α. For the battle of Kursk the values of S can be found in Table 7 (in Appendix).

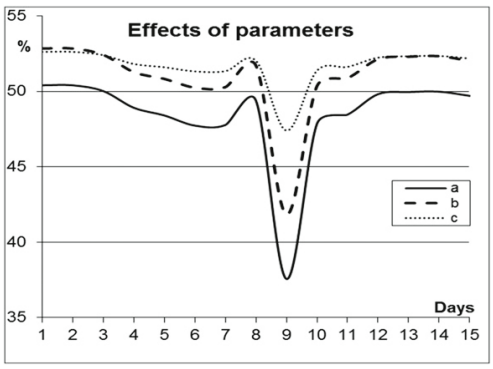

Figure 4. Effects of parameters p_A and α. Parameter values used for curve a: $p_A = 1.0$, $\alpha = 0.1$; curve b: $p_A = 1.1$, $\alpha = 0.1$, and curve c: $p_A = 1.1$, $\alpha = 0.4$.

Without taking into account the decision rules and mechanisms of the battle leaders and decision makers, no theory can produce the probability values for the victory of a battle. The model for the decision boundaries and the corresponding parameters α and β for the opposing forces describes the decision-making process. In our model, the decision boundaries are determined by the initial force sizes and the attrition processes of the two sides. Other effects can be taken into account by adjusting the parameter value of p_A.

The algorithm corresponding to Table 4 (in Appendix) is the following:

$$Set p_A = p_D = 1.0\,(Step2).$$

If $\sigma_B > \sigma_A$ set $\beta = 0$ and adjust α until $P_A = P_D$ else set $\alpha = 0$ and adjust β until $P_A = P_D$ (Step 3).

If the attacker has a clear advantage set a value higher than 1 to p_A ($= p_D$). If the attacker has a clear disadvantage set a value less than 1 to p_A ($= p_D$) (Step 4).

When S is close to 1 the value of p_D is set to p_A. In this section we set $p_D = p_A$ in all calculations because S is close to 1 in all the cases. Secondly, the analysis is done by using the attacker view of combats. We have checked with a spreadsheet application that the defender view gives similar probabilities to the attacker view. In one case, this can be seen in Table 7 (in Appendix) columns P_A and P_D.

Two-sided observations on battle days that include complete data on forces and attrition cannot usually be determined. Most accessible

battle data contains only starting sizes and casualties, sometimes only for one side. Time development of the force strengths on both sides, and casualties on both sides, is available from Kursk (Turkes, 2000), Ardennes (Bracken, 1995) and Inchon-Seoul (Hartley, 1995) combats.

Next, we present a shortened version of the three historical battles mentioned above. In July 1943, the battle of Kursk, the largest tank battle in World War II, took place around the city of Kursk and ended in defeat for the Germans. After the Germans defeat in Stalingrad, the Russians knew they were going to win the war, and the Germans strongly suspected they might lose it. For the battle of Kursk the data covers 15 days of the battle, from July 4, 1943 through July 18, 1943. The days of German attack July 4–July 11 (days 1–8 of the battle) followed the Soviet attack July 12–July 18 (days 9–15 of the battle). More details of actions during the operation can be found in (Turkes, 2000). In the following, Blue denotes the Soviet and Red denotes the Germans.

In the analyses, we regard the days of the battle independent, as if a new combat is occurring day after day during the battle. This describes the daily decision making of the leaders on both sides. We use combat data from Tables 15, 19, and 30 in Turkes (2000) and also the reformatted combat data in Tables 31–34 in Turkes (2000).

The variance is higher and a clear advantage in favor of the Blue forces exists. Consequently, we fix the values of α, p_A, and p_D as following:

$$\alpha = 0,\ p_A = p_D = 1.1. \tag{47}$$

Parameter values α and β affect both the probability level and the shape of vertical direction of the curves. After fixing α or β, we have only one parameter to adjust the exact level of the curves. In normal cases, the values of α and β vary between $(-1, 1)$. However, the theory for the decision boundaries provides no definite limitations. Next, we adjust β until the values P_A and P_D in Equations (38) and (39) are equal. Finding the appropriate values for the parameters is easy with a spreadsheet software. The analyst's task is to calibrate the probability value in one point of the battle, for example, at the beginning or at the end of the battle. This can be accomplished by an expert judgment from the textual description of the battle history. In our case, the Blue force has an advantage and a value greater than one for p_A is justified. In the analysis of the Kursk battle, the same parameter values (47) are used during the calculation.

The choice of the value for p_A shifts the levels of probability in Figure 5. The shape of the curve remains the same. The probability to win the combat is 2–3 percent lower when $p_A = 1.0$. However, the outcome of the combat may be completely different when the probability to win is 2–3 percent lower for every 15 days of the battle. With the value $p_A = 1.1$ most of the days are predicted to be victorious for the Soviet forces while with the value $p_A = 1.0$ most of the days are below the 50 percent probability level.

After determining the parameters of the model with the help of historical descriptions of events and decision making during the operation, the analyses phase follows. In the analysis, the main events and their timing during the battle are compared with the probability curves calculated with the model. In the battle of Kursk, two phases of Red and Blue attacks are clearly shown in Figure 5. Day 9 deviates from the other days as the Blue forces mount their counterattack. By the end of the battle the Blue force retained their initiative and

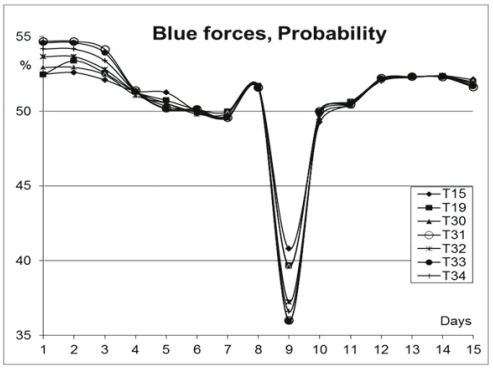

Figure 5. Analysis of the battle of Kursk. Probabilities for the Blue forces to win the battle during the 15 days of the battle. Every day is handled as an individual combat. Data is described in Table 5 (in Appendix). Parameter value $p_A = 1.1$ has been used.

dominated the front until the end of the war. The curves in Figure 5 correspond to manpower and aggregated forces data described in Table 5 (in Appendix). As an example, the probability values for the Blue forces to win the combat at day 9 is shown in the table.

In a complete analysis, probability levels and changes should be explained in connection with the historical information. The model is a tool, which can be used with other methods to understand historical battles, or to predict the outcome of an ongoing battle. The parameter values of the model can be varied during the battle, but in our example, we keep the model simple and make no changes.

Next, we apply the same analysis method to weapon systems from the Kursk battle data. The corresponding probability values for tanks, artillery, and Armored Personnel Carriers (APCs) can be calculated from Equations (38) and (39). The values of β are 0.0, 0.75, 0.7 for tanks, artillery, and APCs, respectively. In these cases, the results have no direct interpretation as probabilities to win the battle. Nevertheless, the results provide another tool for the analysis of the effects of different weapon systems and the battle environment. One conclusion from Figure 7 is that APCs and artillery cancel each other's effects on day 3. Tanks, APCs, and manpower (see Figure 7) strengthen each other's effects on day 9. Casualty numbers of aggregated data T30 and T33 are shown in Figure 6. Casualties of

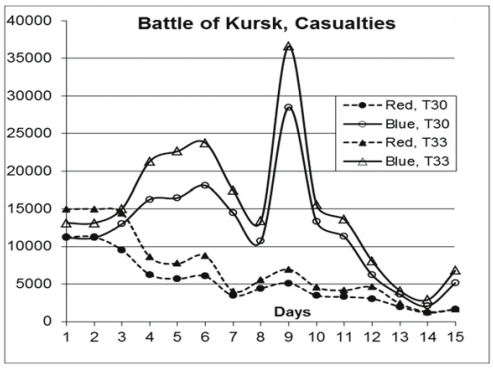

Figure 6. Blue and Red casualties. T30 casualty data is from Table 7 (in Appendix) columns x_c and y_c for Blue and Red casualties correspondingly (Turkes, 2000). T33 casualty data is also from (Turkes, 2000).

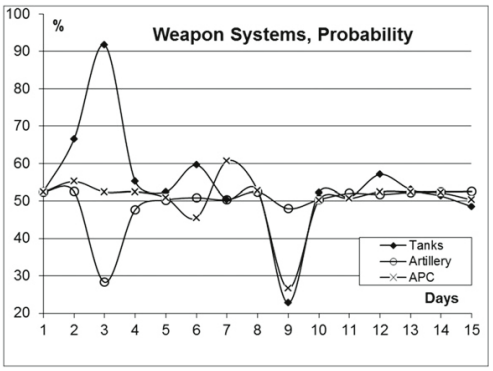

Figure 7. Tanks, artillery, and APC data (Turkes, 2000) from the Kursk battle are analyzed. Parameter value $p_A = 1.1$ has been used.

Blue and Red forces in Figure 6 and Blue probabilities in Figure 5 have a clear correlation, but no obvious functional form can be identified.

The Ardennes Campaign, commonly known as the Battle of the Bulge, caught the Allies by an almost complete surprise. In the following, Blue denotes the Allies (United States and Great Britain) and Red denotes the Germans. However, Red attacks suffered from a number of major weaknesses: personnel quality was weak, transportation networks, and air forces were inadequate, to mention a few. After several days of Red penetrations, the Blue forces rallied to slow down and then to stop the Red attacks. By the beginning about day 4 the Blue forces attacked into the left shoulder of the bulge to relieve beleaguered Blue units. By day 11, the Blue air supremacy was brought to bear on the Red units in the bulge. The Blue forces counterattacked, and two weeks later they restored the front line in the Ardennes. Comparing the Ardennes Campaign description and Figure 8, the main operations can be connected with the probability changes. The actual casualty and reinforcement numbers are used in the analysis with the textual information.

Our third example is the Inchon-Seul campaign of the Korean War. The Battle of Inchon-Seul was an amphibious invasion and the battle of Korean War that resulted in a decisive victory and strategic reversal in favor of the United

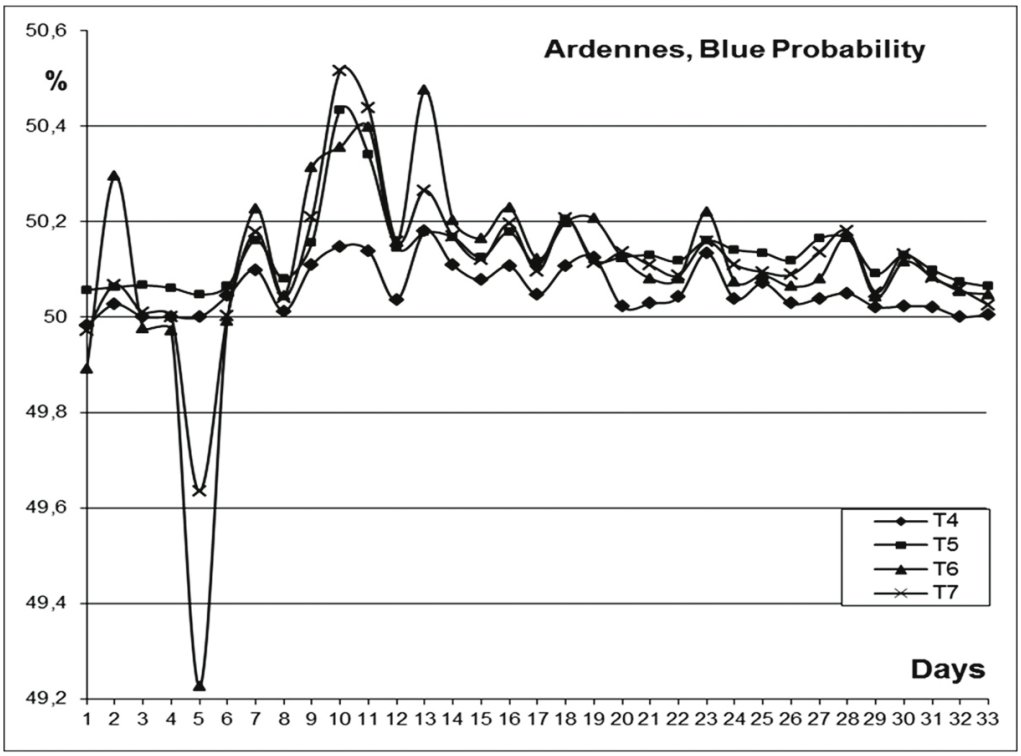

Figure 8. Analysis of the battle of Ardennes. Probabilities for the Blue forces to win the battle during the 33 days of the battle. Every day is handled as an individual combat. Data (Bracken, 1995) is described in Table 6 (in Appendix). Parameter value $p_A = 1.0$ has been used.

Nations. The analysis is shown in Figure 9 for two data sets in Hartley (1995). Blue denotes the South Korean and US forces and Red denotes the North Korean forces. The probabilities are calculated with manpower data and manpower data with possible reinforcements on days 4 and 8. The results are very similar for both data sets, in spite of the fact that reinforcement is as large as about 20 percent on both days. In Hartley (1995), three phases of the campaign have been recognized: days 1–6, days 7–13, and days 14–19. These phases can be seen in Figure 9. The parameter value $p_A = 1.1$ has been used because of the surprise assault of the Blue forces against the Red forces. As a result of added troops, the variance of Blue forces is higher, and without the added troops the variance of Red forces is higher. The decision boundary parameter values used in the first case are $\alpha = 0.63$ and $\beta = 0$ and in the second case $\alpha = 0.0$ and $\beta = 0.15$.

CONCLUSION

A macroscopic combat equation has been derived with stochastic methods for lognormally distributed force sizes. Variables of the equation are the force sizes of the attacker and the defender at the beginning and at the end of the battle.

The model has two parts, the model for stochastic processes and the model for decision boundaries. These can be modeled with alternative processes or functions. In this paper, geometric Brownian motion and linear time-dependent functions are used in the modeling. Decision boundaries have complex dependency

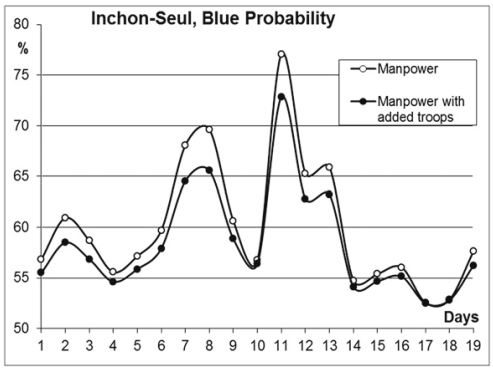

Figure 9. Analysis of the battle of Inchon-Seul. Probabilities for the Blue forces to win the battle during the 19 days of the battle. Every day is handled as an individual combat. Parameter value p_A=1.0 has been used.

on situational conditions and decision rules. In this paper, a simplified model has been presented for the decision boundaries.

Parameters of the model are variances of the attrition processes, slopes of the decision boundaries and advantage measures for the attacker and the defender. All model parameters of the combat equation have a real-world interpretation. Typical parameter values for the slopes of the boundaries and the advantage measures have been presented. The model gives the probability to win a battle for individual battles. The model is more detailed than the Helmbold equation, for example. The theoretical probability values, with appropriate values of the variance parameters, agree with the attrition data. Several earlier models predict the victor of a battle with the same accuracy. In addition, the model in this paper gives the probability value to win a battle. We have also demonstrated with three examples how to analyze the time evolution of a battle. In this analysis, variances of the force sizes are calculated from the data, and variance is not a model parameter. Greater operational insights and predictions of battle outcome can be achieved with the model.

REFERENCES

Ancker, C. J., Jr., and Gafarian, A. V. 1992. *Modern Combat Models, A Critique of Their Foundations*, Operations Research Society of America.

Armstrong, M. J. 2005. A Stochastic Salvo Model for Naval Surface Combat, *Operations Research*, Vol 53, No 5, 830–841.

Bonder, S., and Farrell, R., eds. 1970. *Development of Models for Defense Systems Planning.* Report No SRL 2147 TR 70-2, Systems Research Laboratory, The University of Michigan.

Bracken, J. 1995. Lanchester Models of the Ardennes Campaign. *Warfare Modeling*, Bracken J., Kress M., and Rosenthal, R. E., eds. John Wiley & Sons, 417-435.

Chen, P. S., and Chu, P. 2001. Applying Lanchester's Linear Law to Model the Ardennes Campaign, *Naval Research Logistics*, Vol 48, No 8, 653-661.

Dupuy, T. N. 1985. *Numbers, Predictions &War.* HERO Books.

Fowler, B. W. 1995. De Physica Belli: An Introduction to Lanchesterian Attrition Mechanics Part I. Report. DMSTTIAC SOAR 95-01.

Fricker, R. D., Jr. 1998. Attrition Models of the Ardennes Campaign, *Naval Research Logistics*, Vol 45, No 1, 1-22.

Hartley, D. S., III. 1995. A Mathematical Model of Attrition Data. *Warfare Modeling*, Bracken, J., Kress, M., and Rosenthal, R. E., eds. John Wiley & Sons, 443-465.

Hartley, D. S., III. 2001. *Topics in Operations Research: Predicting Combat Effects*, Institute for Operations Research and the Management Sciences.

Hartley III, D. S., and Helmbold, R. L. 1995. Validating Lanchester's Square Law and Other Attrition Models, *Warfare Modeling.* Bracken, J., Kress, M., and Rosenthal, R. E., eds. John Wiley & Sons, 467-491.

Hausken, K., and Moxnes, J. F. 2002. Stochastic conditional and unconditional warfare, *European Journal of Operations Research*, Vol 140, 61-87.

Hausken, K., and Moxnes, J. F. 2005. Approximations and Empirics for Stochastic War Equations, *Naval Research Logistics*, Vol 52, No 7, 682-700.

Helmbold, R. L. 1989. *Do Battles and Wars Have a Common Relationship Between Casualties and Victory.* Presented at the 57th MORS

Symposium, Fort Leavenworth, Kansas (June 6-8).

Hill, R., McIntyre, G. A, Tighe, T. R., and Bullock, R. K. 2003. Some Experiments with Agent-Based Combat Models, *Military Operations Research*, Vol 8, No 3, 17-28.

Hughes, W. E. 1995. A Salvo Model of Warships in Missile Combat Used to Evaluate Their Staying Power, *Naval Research Logistics*, Vol 42, No 3, 267-289.

Jaiswal, N. K., and Nagabhushana, B. S. 1995. Termination Decision Rules in Combat Attrition Models. *Warfare Modeling*, Bracken, J., Kress, M., and Rosenthal, R. E., eds. John Wiley & Sons, 273-287.

Karlin, S., and Taylor, M. H. 1975. *A First Course in Stochastic Processes Second Edition.* Academic Press.

Kingman, J. F. C. 2002. Stochastic Aspects of Lanchester's Theory of Warfare, *Journal of Applied Probability*, Vol 39, No 3, 455-465.

Lanchester, F. W. 1914. Aircraft in Warfare: The Dawn of the Forth Arm—No. V., The Principle of Concentration, *Engineering*, Vol. 98, 422-423.

Lauren, M. K. 2002. A Fractal-Based Approach to Equations of Attrition, *Military Operations Research*, Vol 7, No 3, 17-29.

Lucas, T. W., and Dinges, J. A. 2004. The Effect of Battle Circumstances of Fitting Lanchester Equations to the Battle of Kursk, *Military Operations Research*, Vol 9, No 2, 17-30.

Mitzenmacher, M. 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions, *Internet Mathematics*, Vol I, No 2, 226-251.

Moffat, J., Smith, J., and Witty, S. 2006. Emergent Behavior: Theory and Experimentation Using the MANA Model, *Journal of Applied Mathematics and Decision Sciences*, 1-13.

Morse, P. M., and Kimball, G. E. 1951. *Methods of Operations Research.* John Wiley & Sons.

Perry, W. 2003. Modeling Knowledge in Combat Models, *Military Operations Research*, Vol 8, No 1, 43-55.

Perry, N. 2011. Applications of Historical Analyses in Combat Modelling. Technical Report DSTO-TR-2643.

Salim, A. S. R., and Hamid, W. M. 2001. A Bayesian Stochastic Formulation of Lanchester Combat Theory, *Military Operations Research*, Vol 6, No 3, 69-76.

Speight, L. R. 2001. Lanchester's Equations and the Structure of the Operational Campaign: Within-Campaign Effects, *Military Operations Research*, Vol 6, No 1, 81-103.

Speight, R. L. 2002. Lanchester's Equations and the Structure of the Operational Campaign: Between-Campaign Effects, *Military Operations Research*, Vol 7, No 2, 15-43.

Taylor, J. G., and Neta, B. 2001. Explicit Analytical Expression for a Lanchesterian Attrition-Rate Coefficient for Bonder and Ferrell's m-Period Target Engagement Policy. Working Paper #5. DTRA Project.

Turkes, T. 2000. *Fitting Lanchester and Other Equations to the Battle of Kursk Data.* Master's Thesis. Department of Operations Research, Naval Postgraduate School, Monterey, California.

Willard, D. 1962. Lanchester as Force in History: An Analysis of Land Battles of the Years 1618–1905. No AD297375, DTIC, Alexandria, Virginia.

APPENDIX

Table 4. The main steps of the analysis.

Step 1	Data collection.	Data on forces and attrition is used as an input. Observations on important events and actions on battle days are used for the analysis.
Step 2	Preliminary analysis.	In preliminary analysis the information is analyzed to find out a possible superiority of one side of the conflict at the beginning of the combat.
Step 3	Setting the decision boundary parameters.	The decision boundary parameters α and β are adjusted. First, in a spreadsheet application the value of the decision boundary parameter (α or β) is set to 0 for the force having higher variance value and the second boundary parameter value is adjusted until the attacker view P_A and the defender view P_D probabilities are equal in Equations (38) and (39). (If the adjusted parameter value gets abnormal values greater than 1 or less than -1, values of the first parameter other than 0 between -1 and 1 can be tried. At this phase, the information about the battle and graphical curves of the decision boundaries may be helpful for the analyst. No abnormal cases like this exist in manpower or combat power data used in this study.)
Step 4	Setting the value of the advance parameter p_A.	Using the results of the preliminary analysis, the advance parameter p_A is fixed. The default value for p_A is 1.0. If the Blue forces (or the attacker) has a clear advantage, the advance parameter is set to a higher value.
Step 5	Final analysis.	In the final analysis the information collected in Step 1 and the probability curves are compared. Probability levels and changes on battle days are linked with important actions of the battle.
Step 6	Documenting.	Conclusions and visual presentations are collected as a report. Graphical curves of casualty rates and weapon system changes as a function of time improve the understanding of the battle developments.

Table 5. Description of the data of Figure 5. Also model parameter values of β and calculated day 9 probabilities are shown.

Table in Turkes (2000)	Description of the data	β	Day 9 (%)
T15	Combat manpower data. Casualties are killed, wounded, captured/missing in action, and disease and non-battle injuries.	0.82	41
T19	Data on aggregated forces. Forces are combat manpower, APCs, artillery and tanks which are weighted by 1, 5, 20, and 40, respectively.	0.82	40
T30	Data on aggregated forces. Forces are combat manpower, APCs, artillery, tanks and number of ground-attack role sorties which are weighted by 1, 5, 20, 40, and 30 respectively.	0.39	40
T31–34	Data on aggregated forces reformatted using the Bracken method presented in (Turkes, 2000). Forces are combat manpower, APCs, artillery and tanks, which are weighted by 1, 5, 15, and 20 or 1, 5, 15, 20 or 1, 5, 30 and 40 or 1, 5, 20, and 30 in Tables 31–34 in Turkes (2000), respectively.	0.87	36–37

Table 6. Description of the data of Figure 8. Also model parameter values of β are shown.

Table in Bracken (1995)	Description of the data	β
T4	Data on combat manpower (infantry, armor, and artillery).	0.000
T5	Data on combat forces.	0.198
T6	Data on total manpower	0.201
T7	Data on total forces	0.200

Table 7. Data from the battle of Kursk (T30 in Table 5) and the results calculated from the model of this paper. Parameters of the model are α, p_A, β, and p_D. Variances σ_A^2 and σ_B^2 are variances of columns $\ln(x_0 - x_c)$ and $\ln(y_0 - y_c)$ for the attacker and the defender correspondingly. Casualties are denoted x_c and y_c. Corresponding equations in the text are also shown in the table. The table is a copy of the spreadsheet application used in the analysis of this study.

T30	α	p_A	β	p_D	σ_A^2	σ_B^2
	0	1.10	0.39	1.10	**0.0089**	**0.0027**

Equations in the text:							*Eq. 28*	*Eq. 29*	*Eq. 34*	*Eq. 38*	*Eq. 30*	*Eq. 40*	*Eq. 39*
Days	x_0	x_c	y_0	y_c	$\ln(x_0 - x_c)$	$\ln(y_0 - y_c)$	S	m_A	A	P_A	m_D	D	P_D
1	604353	11167	431671	11257	13.293	12.949	1.0078	−1.751	0.984	0.529	3.489	1.010	0.527
2	604353	11167	431671	11257	13.293	12.949	1.0078	−1.751	0.984	0.529	3.489	1.010	0.527
3	594159	12993	404945	9532	13.273	12.888	1.0017	−0.385	0.996	0.524	0.768	1.002	0.524
4	579175	16266	404055	6249	13.241	12.894	0.9872	2.905	1.027	0.511	−5.791	0.984	0.514
5	565402	16472	415304	5702	13.216	12.923	0.9844	3.545	1.033	0.505	−7.066	0.980	0.508
6	542712	18071	406024	6043	13.170	12.899	0.9813	4.249	1.040	0.498	−8.470	0.976	0.501
7	527893	14445	382404	3450	13.149	12.845	0.9815	4.207	1.040	0.499	−8.386	0.976	0.502
8	518016	10754	389340	4415	13.137	12.861	0.9905	2.156	1.020	0.516	−4.298	0.988	0.519
9	498123	28492	375765	5112	13.060	12.823	0.9558	10.18	1.097	0.396	−20.29	0.943	0.385
10	487961	13302	375759	3491	13.070	12.827	0.9819	4.122	1.039	0.499	−8.217	0.977	0.503
11	480724	11323	394230	3290	13.059	12.876	0.9847	3.481	1.033	0.506	−6.938	0.980	0.509
12	474229	6201	373752	3047	13.056	12.823	0.9950	1.121	1.010	0.521	−2.234	0.994	0.523
13	482881	3600	367286	1975	13.080	12.809	0.9979	0.471	1.004	0.523	−0.939	0.997	0.524
14	471266	2067	363905	1174	13.059	12.801	0.9988	0.262	1.002	0.523	−0.523	0.999	0.524
15	469253	5160	360820	1639	13.048	12.792	0.9935	1.465	1.014	0.52	−2.920	0.992	0.522

V

Decision Boundaries Used to Model Probability of Victory and Duration of Combats

Vesa Kuikka

Journal of Applied Operational Research, 9(1), 2017, pp. 67-81

Available at http://www.orlabanalytics.ca/jaor/archive/v9/n1/jaorv9n1p67.pdf

Correction: "Equation (5) is a form of Equation 4.18 in (Taylor & Karlin, 1998)" on page 72 should be "Equation (5) is a form of Equation 4.19 in (Taylor & Karlin, 1998)"

Journal of Applied Operational Research (2017) Vol. **9**, No. 1, 67–81 ISSN 1735-8523 (Print), ISSN 1927-0089 (Online)

www.orlabanalytics.ca

Decision boundaries used to model probability of victory and duration of combats

Vesa Kuikka [1,*]
[1] Finnish Defence Research Agency, Riihimäki, Finland

Received 30 October 2016
Accepted 15 December 2017

Keywords:
Attrition combat data
Combat modelling
Decision boundary
Duration of a battle
Geometric brownian motion
Probabilistic modelling

Abstract—Geometric Brownian motion with decision boundaries is used in modelling combat effects such as probability of winning a battle and duration of a battle. We verify that it is possible to construct a model and analytical mathematical formulas for these quantities. Variables of the model are force sizes at the beginning and at end of a battle, and decision boundaries for the two opposing forces. The model with different model parameters is compared with empirical attrition data. New results can be uncovered when outcomes and expected durations of empirical battles are studied with the model. We conclude that battles, where the attacker is superior to the defender, and battles with comparable force strengths on the two opposing sides, are described with different model parameter values. This can indicate an extra advantage for attackers that cannot be explained by other factors of the model. We show how asymmetrical decision boundaries are used to model this kind of effects. These general conclusions are made on an aggregate level from the modelling results of individual battles.

Published online 31 December 2017

Introduction

Understanding relationships between the outcome of a battle and factors like force sizes of opposing sides is a fundamental question in military studies. For an army leader to be able to predict combat effects more accurately would be vitally important. Any improvements in tools, methods or understanding would be most welcome. Historically force size is the most important factor but force quality, tactics, casualties, duration, surprise, advance rates, and victory are also important (Hartley, 2001). In modern warfare, technological systems, not only weaponry but also communication and information systems have an increasing importance. The role of information has increased substantially in the last few decades by UAVs and other electromechanical means. Changes are likely to continue, especially in asymmetrical conflicts. Tactics has also changed from historical battles, distributed and networked operation mode has been adopted by many armed forces.

Methods used to study combat effects are numerous and cover many fields of military sciences, information analysis, mathematics, psychology, and social sciences. Much of the work has been heuristic in the sense that results are not presented in quantitative terms. Quantitative methods are usually mathematical formulas or simulation methods. One line in the long history of studying combats is trying to write down an analytic combat equation, as simple as possible, describing combat effects, for example predicting the winner of a battle. However, only few closed form equations exist today. One may argue that no simple formula can describe complex relationships in warfare, and modern warfare is becoming even more complex. In our view, this does not exclude the need to understand the big picture of the complex system of combats. If there are common features in all combats or in in specific types of combats, these features should be recognized and studied with appropriate methods. Secondly, mathematical models and results from the analysis may give a benchmark for comparing and characterizing combats. This would give a basis for more detailed and comprehensive models. Knowledge about which are the most important factors and what are the interrelationships between them is the basic work necessary before a more detailed theory can be developed. Comparing predictions of models and empirical observations provide information about assumptions of the models. Modelling should be realistic enough to enable correct conclusions in these kinds of investigations.

Quantitative equations should be understandable and derived and expressed with real world quantitative terms. In this respect, some combat models are phenomenological and others are based on exactly defined quantities. Examples for the latter are the Lanchester equations (Lanchester, 1914). The Lanchester equations are expressed with force sizes and attrition rates. Examples of phenomenological formulations are the Helmbold relationship (Helmbold, 1989) and the Willard

* *Correspondence*: Vesa Kuikka, Finnish Defence Research Agency,
PO BOX 10. Tykkikentäntie 1. FI-11311 Riihimäki, Finland
E-mail: vesa.kuikka@mil.fi

equation (Willard, 1962). They use force sizes as input values but involve some phenomenological parameters without real word interpretation. Typically these parameters are fitted with empirical data. Models having both types of variables, phenomenological and empirical, are also common.

Both types of methods are useful but our goal in this paper is to give a quantitative interpretation or explanation to all the expressions or at least give guidelines for future studies how to construct a more detailded quantitative model. Decision boundary is an example of a concept that can be considered as a phenomenological parameter but in theory it can be defined exactly, for example, based on enquiries among combat leaders. There is a real word meaning behind the concept but the problem is how to define it exactly and how to relate it with other relevant quantities of the problem.

A number of historical battle data samples have been aggregated and used in (Hartley, 2001). The aggregate database covers a wide range of force ratios, and while emphasizing 20th Century battles, has reasonable coverage back to 1600. The database has numerical values for force sizes at the beginning of a battle and casualties at the end of a battle for both sides of conflicts. In addition, the empirical data includes durations for each battle. The database of almost 600 lines indicates the attacker and provides the information of the winner of each battle. The empirical data used in this study is from (Hartley, 2001).

Our concern in this paper is limited in closed form relationships between force sizes, decision rules, probabilities and durations of battles. The model of this paper is stochastic in the sense that a particular stochastic process, geometric Brownian motion, is used as a model for attrition process. The results can be expressed analytically and this is why no simulation is needed. The model is based mostly on manpower size and decision boundaries. A geometric Brownian motion (Taylor & Karlin, 1998) is defined with two parameters, variance and drift. Variance describes stochastic variability and drift describes attrition rate of the stochastic process.

Other factors mentioned earlier may be incorporated in the model by constructing models for decision boundaries or force sizes with these factors. Technological factors such as transport and logistics can be modelled with decision boundaries. Supporting personnel taking care of dead and wounded can be modelled as manpower. After all, it is a modelling problem to take these factors into account in a special case. One possibility is to consider the manpower and the weaponry (Dupuy, 1985) actually taking part in engagements as force strength and other elements as factors changing decision boundaries. In this case, reinforcement is modelled as a part of manpower and reserve forces have effects on decision boundaries. The model of this paper is not primarily designed for studying asymmetrical battles or conflicts when one side is a non-traditional force where the force numbers are unknown.

Associated with the question of modelling details is the fact that quality of the data of historical battles is not high. Different conventions in documenting force strengths and durations of battles have been used. Defining the winner of a battle is not clear in all situations but basically surrendering, not just retreating, is considered as losing a battle. In this study, we take the historical data as it stands (Harley, 2001).

In many battles, heavy fighting takes place just before the end of a battle. This can cause a bias in modelling if the drift values of real attrition processes of attackers and defenders are modelled individually. In this paper, we use a definition for the attrition process and its drift that is different from our earlier paper (Kuikka, 2015). Here, we define the stochastic process as the difference of attacker's attrition process and defender's attrition process. The drift of this process is defined correspondingly as the difference of attacker's drift and defenders drift. In part, different conventions of writing down the attrition information of an attacker and a defender cancel each other.

In the next section, we present related work considering analytical macroscopic combat equations. The model of this paper and related work of (Hartley, 2001) have comparable agreement with empirical data. No clear attacker's advantage has been found in earlier studies (this can also be a consequence of low quality of empirical data). Next, background from earlier work of the author of this paper (Kuikka, 2015) introduces basic concepts. The mathematical formulas for geometric Brownian motion are presented and they can also be found in (Taylor & Karlin, 1998). In this paper a new model for decision boundaries is defined that enables us to predict durations of battles. We can compare, on an aggregate level, theoretical results of probabilities to win battles and expected durations with empirical data.

The empirical attrition data (Hartley, 2001) has no information about variances or decision boundaries of battles. In the model, low variance values or wide decision boundaries can both explain long durations of battles. We show that decision boundaries have a greater effect than different variance values on probabilities and expected durations in battles where attrition rates of the opposing sides are very different. Variance are found to be important in modelling phenomena of even battles and decision boundaries are more important in modelling more extreme battles. Both of these factors can have effects especially on individual battles. In addition, we use decision boundaries for modelling asymmetrical effects between attackers and defenders. The results show that attackers have an extra advantage which is difficult to explain by other factors of the model, i.e. force strengths and variances of attrition processes.

Related work

Combat equations have been studied more than hundred years. The Lanchester equations (Lanchester, 1914) are probably the most known and used deterministic combat equations. The Lanchester equations are solutions to differential equations describing attrition of two fighting troops.

Other closed form combats equations are the Helmbold relationship (Helmbold, 1989; Hartley, 2001) and the Willard equation (Willard, 1962). The Helmbold relation is as follows:

$$\ln\left(\frac{x_0^{\,2}-x_u^{\,2}}{y_0^{\,2}-y_d^{\,2}}\right)=c_1\ln\left(\frac{x_0}{y_0}\right)+c_2, \tag{1}$$

where x_0 and y_0 are the force sizes of the attacker and the defender at the beginning of a battle. And x_u and y_d are the force sizes of the attacker and the defender at end of a battle. In Equation (1) c_1 and c_2 are constants. We use the same notation in Equation (1) as in other equations of this paper. Subscripts *u* and *d* denote "up" and "down" to describe upper and lower decision boundaries, but in the Helmbold relationship the meaning is more like "attacker" and "defender".

Extensive research with empirical combat attrition data has been conducted in (Hartley, 2001). The work is based on the Helmbold relation where several factors including force sizes, environmental variables, and human factors have been used as explanatory variables to predict the winner of a battle. The work is mostly heuristic and no modelling from first principles or based on quantitative real world quantities has been done. This is a practical approach taking into account the fact that building a complete theory with so many variables is a difficult task. Some factors can be incorporated in models by including them in force sizes, for example one tank equated as a specific number of soldiers (Dupuy, 1985). Evaluating equivalent numbers of systems, weaponry, and fighters is a complex problem. One method of comparing different sources of military capability and determining equivalent numbers of soldiers and system units is optimizing total capability values (Kuikka, 2016).

Basic Lanchester equations are deterministic. Simulated results based on the Lanchester models have been presented in the literature. Also, stochastic versions of the Lanchester equations and probabilistic models based on Markov chains have been proposed. As a model becomes more complex, solving it analytically becomes intractable or computationally expensive. Approximate solutions have been proposed for computationally intensive problems, such as optimal resource allocation and analysis of asymmetric forces like snipers and fighter aircrafts (Kim *et al*., 2017; Lappi *et al*. 2012). Most of these models calculate force sizes and attrition rates while probabilities to win battles and durations of battles are not considered. Deterministic and stochastic models may be extended with the concept of decision boundaries in order to compute these quantities. This line of work is a subject of possible future research. More comprehensive literature review including also stochastic and simulation methods can be found in (Hartley, 2001; Kuikka, 2015).

Background from earlier work

Our earlier paper (Kuikka, 2015) presented a combat model based on geometric Brownian motion with moving decision boundaries. In a way, methods of this paper and earlier work (Kuikka, 2015) can be considered as two manners of approaching the same problem. Attrition processes were modelled as actual force sizes separately for the two opposing forces. Decision boundaries were modelled as linear functions of time. The slope for attacker's decision boundary was determined by the slope of defender's force size and vice versa. This model has a property of descending decision boundaries when opposing sides are losing manpower. The probability of winning a battle is expressed in the following form (Kuikka, 2015).

$$P=\frac{\left(S^{\alpha-1}\right)^{-|m|}-\left(2S^{\alpha-1}-1\right)^{-|m|}}{1-\left(2S^{\alpha-1}\right)^{-|m|}}\text{, where} \tag{2}$$

$$S=\frac{x_u y_0}{x_0 y_d}\text{ and} \tag{3}$$

$$m = 1 - \frac{2}{\sigma^2}(1-\alpha)\ln(S). \tag{4}$$

Combat equation (2) is different from combat equations to be presented later in this paper because of different process and decision boundary specifications. In our earlier paper the process and its drift were modelled as the process and its drift of the attrition process itself.

In Equations (2) and (4) α is a model parameter for adjusting how steep is the decision boundary and σ is volatility (σ^2 is variance in a time unit) of the attrition process. The two bars around m denote absolute value. In Equation (3) $x_0, y_0, x_u,$ and y_d are the initial attacker force strength, initial defender force strength, final attacker force strength, and final defender force strength. To be precise, Equation (2) describes the attacker view of a battle. The corresponding defender view looks much the same and the equation can be found in (Kuikka, 2015) as well as the derivation of the formulas. We will use the quantity S of Equation (3) throughout in this paper as an independent variable. The quantity S itself is a good predictor of battle outcomes. If S is less than one the attacker's probability to win is less than 50 % and if S is greater than one the probability is greater than 50 %. However, Equation (3) does not provide numerical probability values as a function of force sizes.

One nice feature of the model is that the equations describing the attacker view and the defender view give the same results if variances on both sides are equal. This can be regarded as a consistency requirement for the model. If the variances are not equal α-parameters must be different because still the attacker and the defender views should give similar results. This procedure has been explained in (Kuikka, 2015) with examples of the few cases where the variances are actually known from historical data.

In this paper decision boundaries are constants as a function of time. The decision boundary definition is related with the definitions of attrition processes and drift terms. In this paper, drift is defined as the difference of the drift of attacker's attrition process and the drift of defender's attrition process. A stochastic process describing the combined attrition process of a battle is approximated with a geometric Brownian process. The new modelling enables predicting both probabilities and durations of battles, not just probabilities as in (Kuikka, 2015). Comparing both of these results, probabilities and durations, with empirical data, can reveal new properties of combat effects. As attrition processes are modelled with geometric Brownian processes, the drift parameter has the standard mathematical definition provided in the literature (https://en.wikipedia.org/wiki/Geometric_Brownian_motion, accessed 25 December 2017; Karlin & Taylor, 1974).

One goal of this paper is to extend the model to predict combat durations and to compare predicted durations with the observed empirical durations of battles. A deficiency of decision boundaries used in (Kuikka, 2015) is that, when S is close to one in Equation (3), durations became small. This contradicts the observed durations of historical combat data – even though a considerable portion of data points are near (1,0) in (S,T)-coordinates (see Figure 2 later in the text).

We show how the modelling in (Kuikka, 2015) can be carried out with different decision boundaries. A better combat model could be developed with better decision boundary sub-modelling. The problem of constructing more realistic models for attrition processes and decision boundaries still remains a subject for future studies.

Probability to win a battle and duration of a battle

We present a variant of our earlier combat equation with new results concerning durations of battles. Constant decision boundaries are stopping boundaries for the combatant party to end a battle. Two decision boundaries for each side are used in the model, for winning and losing a battle. Attrition processes are approximated with geometric Brownian motion describing the difference of attrition processes of the two sides of a conflict. Initial values of the processes are the initial force strengths of attackers and defenders.

The input data for the model are the initial force sizes of the two opposing sides and the corresponding force sizes at the end of a battle. We denote the attacker by x and the defender by y because the empirical combat attrition data is available in this form (Hartley, 2001). The theory allows different parameter values for attackers and defenders. The parameters of the model are variances (or mean deviations) of attrition processes and decision boundaries. At the most, two variances and four decision boundaries may exist in the model. On an aggregate level, examining only one or two variance parameter values and one or two values for decision boundaries may be sufficient. The reason is that the empirical data has similarities between attackers and defenders in the view of our model.

Earlier studies (Willard, 1962; Helmbold, 1989; Hartley, 2001) don't provide direct evidence of an extra advantage for defenders or attackers that cannot be explained by force strengths on an aggregate level. Attackers more often win battles but so far the models have suggested that the superiority of the force sizes can explain the outcome. In this context, we

mean by superiority that defender's percentage casualties are higher when compared with attacker's percentage casualties ($S > 1$ in Equation (3)). On the other hand, in many cases the defender is well protected and has a good situational awareness of its own territories. In subsets of the available empirical data, these kinds of advantage factors may be needed in modelling. Individual battles are very diverse and the model should be used with different appropriate parameters for the attacker and the defender when considering an individual battle. In individual battles, surprise can be modelled with different decision boundaries and different variance values for the opposing sides. In summary, conclusions about possible extra advantage factors for attackers or defenders are dependent on details of the used model and specifically how realistic the model is.

We can study the problem on an aggregate level and at individual combat level. The available empirical data has no variance information which should be known or evaluated for considering the results at combat level. In this paper, we model attrition processes at combat level but investigate the results on an aggregate level. A typical variance value for all the battles in the data set is assumed. On an aggregate level, we present typical parameter values with justification for their choices. We compare the results of the model with empirical data.

Notations

x_0	Initial force size of attacker
y_0	Initial force size of defender
x_u	Final force size of attacker, $x(T)$
y_d	Final force size of defender, $y(T)$
A, D	A = Attacker, D = Defender
μ	Drift value in one day of attrition process (μ_A attacker and μ_D defender), $\mu = \mu_A - \mu_D$ (https://en.wikipedia.org/wiki/Rate_of_return#Logarithmic, accessed 21 December 2017)
σ^2	Variance of attrition process in one day (σ_A^2 attacker and σ_D^2 defender), $\sigma^2 = \sigma_A^2 - 2Cov(A,D) + \sigma_D^2$, where $Cov(A,D)$ is covariance between attacker and defender attrition processes (https://en.wikipedia.org/wiki/Covariance, accessed 21 December 2017)
T	Duration of a battle $(T_A = T_D)$, $E(T)$ expected value of duration
P_A, P_D	Probability to win a battle, $P_A = 1 - P_D$
Δ_{xu}, Δ_{xd}	Upper decision boundary is Δ_{xu} higher than x_0 for attacker, i.e. $X_u = x_0 + \Delta_{xu}$. Lower decision boundary is Δ_{xd} lower than x_0 for attacker, i.e. $X_d = x_0 - \Delta_{xd}$.
Δ_{yu}, Δ_{yd}	Upper decision boundary is Δ_{yu} higher than y_0 for defender, i.e. $Y_u = y_0 + \Delta_{yu}$. Lower decision boundary is Δ_{yd} lower than y_0 for defender, i.e. $Y_d = y_0 - \Delta_{yd}$.
$\delta_{xu}, \delta_{xd}, \delta_{yu}, \delta_{yd}$	Used in the simple model for decision boundaries: $X_u = x_0 + \delta_{xu} x_0 = (1 + \delta_{xu}) x_0$, $X_d = (1 - \delta_{xd}) x_0$

We develop the model in two phases. Before that, the general formulation of the theory is presented. We assume that the attrition processes can be described as geometric Brownian motions (https://en.wikipedia.org/wiki/Geometric_Brownian_motion, accessed 30 October 2016). Two important attributes of the assumption are the stochasticity of the process and the proportionality of the attrition to the force sizes. In the literature (Perry, 2011) has studied geometric Brownian motion in the context of combat modelling. We adopt two results from the literature, the probability to hit the upper stopping boundary and the duration of the process to hit either upper or lower stopping boundaries (Willmott, 2000; http://marcoagd.usuarios.rdc.puc-rio.br/hittingt.html, accessed 30 October 2016). For the purpose of our model these provide the probability to win a battle and the duration of a battle. We use the notations in above. We choose the attacker point of view; the defender view gives comparable results. The relationship between the probabilities holds that the attacker's probability to win equals the defender's probability to lose. Upper and lower decision boundaries for the attacker are denoted by X_u and X_d. The probability for an attacker to win a battle is

$$P = \frac{(x_0)^m - (X_d)^m}{(X_u)^m - (X_d)^m} = \frac{\left(\frac{x_0}{X_u}\right)^m \left(1 - \left(\frac{X_d}{x_0}\right)^m\right)}{1 - \left(\frac{X_d}{X_u}\right)^m}, \text{ and} \tag{5}$$

the duration of a battle is

$$E(T) = \frac{1}{\frac{\sigma^2}{2} m} \left(\ln\left(\frac{x_0}{X_d}\right) - \frac{1 - \left(\frac{x_0}{X_d}\right)^m}{1 - \left(\frac{X_u}{X_d}\right)^m} \ln\left(\frac{X_u}{X_d}\right) \right), \tag{6}$$

where $m = 1 - 2\frac{\mu}{\sigma^2}$. (7)

Equation (5) is a form of Equation 4.18 in (Taylor & Karlin, 1998) and Equation (6) is from Section 3 Paragraph "Expected First Hitting Time for Either Upper or Lower Boundaries (Geometric Brownian Motion)" in (http://marcoagd. usuarios .rdc. puc-rio.br/hittingt.html). Several sub-models can be created from Equations (5-7) by specifying different models for decision boundaries and variants of attrition processes. Equations (5-7) assume that the stochastic processes describing attrition processes obey geometric Brownian motion. Analytical mathematical results are available for Brownian motion and geometric Brownian motion but for other processes numerical simulations should be used in most cases.

Real attrition processes are descending (with reinforcement ascending processes are possible) and this aspect should be considered at the same time with the specification of decision boundaries. In this paper we study a stochastic process defined as a difference between attacker and defender attrition processes with the drift definition $\mu = \mu_A - \mu_D = \ln(S)$, where S is defined in Equation (3). This definition allows constant decision boundaries: $X_d \le x_0 \le Y_u$ and $Y_d \le y_0 \le Y_u$. The end of a battle occurs when attrition processes of the opposing sides hit upper and lower stopping boundaries at the same time *T*. In the model the winner hits its upper boundary and the loser hits its lower boundary.

Because the stochastic process to be investigated is a difference between two real attrition processes, the variance value is $\sigma^2 = \sigma_A^{\ 2} - 2Cov(A,D) + \sigma_D^{\ 2}$, where $Cov(A,D)$ is covariance between the attacker attrition process and the defender attrition process. Variance is less than the sum of the variances of the two attrition processes. Real attrition processes are positively correlated because casualties of the two sides of a battle are stochastically related with each other.

In the following sections we present two sub-models for the (constant) decision boundaries with the help of available information about attrition processes. In the first model we define the decision boundaries simply as $X_u = x_0 + \delta_{xu} x_0$ and $X_d = x_0 - \delta_{xd} x_0$ for the attacker, and $Y_u = y_0 + \delta_{yu} y_0$ and $Y_d = y_0 - \delta_{yu} y_0$ for the defender. In the second model we define the decision boundaries as $X_u = x_0 + d_{xu}(x_0 - x_u)$ and $X_d = x_0 - d_{du}(x_0 - x_u)$ for the attacker, and $Y_u = x_0 + d_{yu}(y_0 - y_d)$ and $Y_d = y_0 - d_{yd}(y_0 - y_d)$ for the defender. Initial force strengths are denoted by x_0 for the attacker and y_0 for the defender. Force strengths at the end of the battle are x_u for the attacker and y_d for the defender. Numerical values for the force strengths are available from the empirical data. In the formulas the constants $\delta_{xu}, \delta_{xd}, \delta_{yu}, \delta_{yd}, d_{xu}, d_{xd}, d_{yu}$ and d_{yd} are parameters of the models. If the attacker wins X_d and Y_u describe the boundaries that are not directly observed. . If the defender wins X_u and Y_d describe the boundaries that are not directly observed.

There is a relationship between decision boundaries and variances: low variances or wide decision boundaries and high variances or narrow decision boundaries provide similar results for expected durations. Numerical values for decision boundaries and variances are not available from the empirical data (Hartley, 2001). In the second model we vary decision boundary values and use typical variance values to compare the results of the model and the empirical data. However, in the next section we experiment with different variance values and decision boundaries to get some insight about their typical range.

Model 1

In this section, we present a simplified model to demonstrate the effects of different parameter values. We show how the duration of a battle changes with variance values, with different decision boundaries and with asymmetrical decision boundaries. In this introductory model decision boundaries are δx_0 above and below the initial force size value x_0, where δ is a constant. If $\delta = \delta_{xu} = \delta_{xd}$ $\left(\Delta_{xu} = x_0 + \delta x_0, \Delta_{xd} = x_0 - \delta x_0\right)$ and the upper decision boundary is 10 % above x_0 and the lower decision boundary is 10 % below x_0, we get from Equations (5) and (6) for the probability to win a battle

$$P = \frac{\left(\frac{x_0}{x_0 + \delta x_0}\right)^m \left(1 - \left(\frac{x_0 - \delta x_0}{x_0}\right)^m\right)}{1 - \left(\frac{x_0 - \delta x_0}{x_0 + \delta x_0}\right)^m} = \frac{\left(\frac{1}{1+\delta}\right)^m \left(1 - (1-\delta)^m\right)}{1 - \left(\frac{1-\delta}{1+\delta}\right)^m} = \frac{\left(\frac{1}{1.1}\right)^m \left(1 - (0.9)^m\right)}{1 - \left(\frac{0.9}{1.1}\right)^m}, \tag{8}$$

and for expected duration

$$E(T) = \frac{1}{\frac{\sigma^2}{2} m} \left(\ln\left(\frac{1}{1-\delta}\right) - \frac{1 - \left(\frac{1}{1-\delta}\right)^m}{1 - \left(\frac{1+\delta}{1-\delta}\right)^m} \ln\left(\frac{1+\delta}{1-\delta}\right) \right) =$$

$$\frac{1}{\frac{\sigma^2}{2} m} \left(-\ln(0.9) - \frac{1 - \left(\frac{1}{0.9}\right)^m}{1 - \left(\frac{1.1}{0.9}\right)^m} \ln\left(\frac{1.1}{0.9}\right) \right), \quad \text{where} \tag{9}$$

$$m = 1 - 2\frac{\ln(S)}{\sigma^2} \text{ and } S = \frac{x_u y_0}{x_0 y_d} = \frac{(x_0 + \delta x_0) y_0}{x_0 y_d} = \frac{1.1 y_0}{y_d}.$$

With Equations (8) and (9) we can show some typical effects of the parameters σ^2 and δ. Equations (8) and (9) are for symmetrical decision boundaries. Equations for asymmetrical decision boundaries can be expressed in a comparable manner. The effects on the expected duration value of a battle are investigated as a function of S in Equation (9) (equivalently as a function of y_0 / y_d due to the definition of m). In Figure 1A four curves with σ^2 = 1/100, 1/200, 1/300 and 1/400 are shown. Near $S = 1$ ($y_d = 1.1 y_0$) values of $E(T)$ are higher for lower values of σ^2. This is a direct consequence of the functional form of Equation (9). This effect prevails only when σ^2 is approximately between 0.8 and 1.25. Outside this interval $E(T)$ is almost independent of variance value. When $S = 1$ the value of $E(T)$ is proportional to $1/\sigma^2$. The expected duration values of a battle are 1, 2, 3, and, 4 days for the variance values mentioned earlier. The interpretation is obvious: higher variation means heavier fighting and shorter duration of a battle.

In Figure 1B a similar comparison with wider decision boundaries is shown. The variance value of σ^2 = 0.01 is used and the effects of decision boundaries symmetrically 10 %, 15 % and 20 % above and below the initial value of force size x_0 are shown. With wider decision boundaries $E(T)$ is higher. This occurs with all values of S in contrast to the curves shown in Figure 1A. The interpretation is that the adversary is more resilient and is not ready to surrender. The battle takes a longer time and the higher expectation time for duration follows.

In Figure 1C an interesting phenomenon can be seen when the decision boundaries are not symmetrical. In this case the upper decision boundary is higher than the corresponding symmetrical upper decision boundary with respect to x_0 would be. In Figure 1C $\delta_{xu} = 0.1$ is the symmetrical case and δ_{xu} = 0.2, 0.3 and 0.4 are asymmetrical results. The value for the lower boundary $\delta_{xd} = 0.1$ is used in all the cases. The consequence of non-symmetricity is asymmetrical behavior also in $E(T)$. With high values of S the expected duration value of a battle is higher when compared with the symmetrical case. It is noticeable that effects with low values of S are negligible. The interpretation is that when $S > 1$ the attacker is superior

and probably closer to the victory, that is, closer to the upper decision boundary and changes on the upper boundary are more important than changes on the lower boundary far away down. The situation is reversed when $S < 1$.

The behavior of the expected duration values in the three situations of Figure 1 can be used in reasoning whether a consistent model of combat can be constructed which have force sizes, variances, decision boundaries and durations as explanatory factors or predicted values.

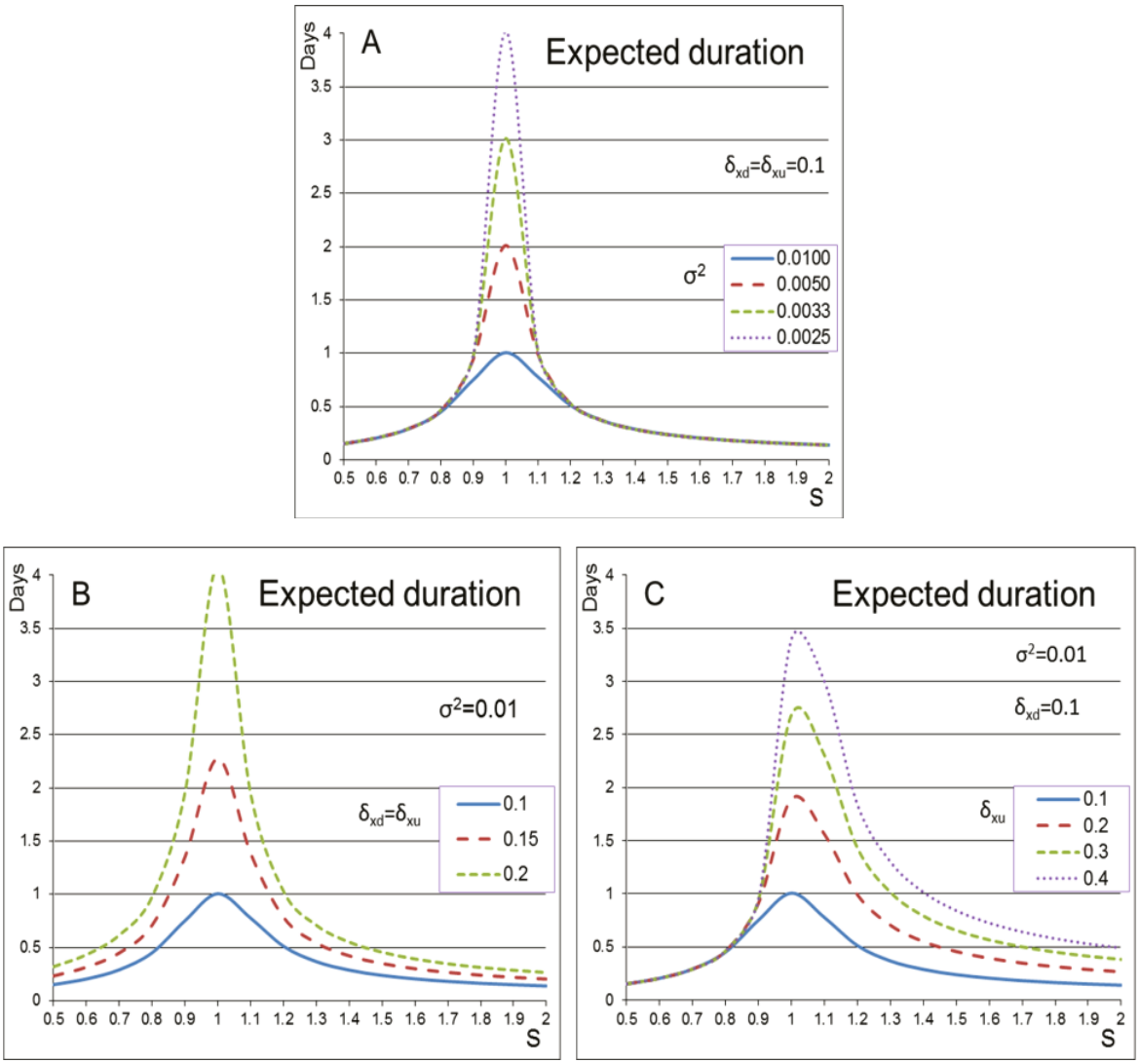

Figure 1. Effects on expected combat durations from Equation (9). A) Different variances, B) Different decision boundaries, C) Asymmetrical decision boundaries.

Model 2

We proceed with a model of more detailed definition for decision boundaries. Decision boundaries depending on a particular condition of a battle may predict outcomes of battles better. From Equations (5-6) with $\mu = \ln(S)$ the probability to win a battle and expected duration can be expressed as

$$P=\frac{(x_0)^{1-\frac{2\ln(S)}{\sigma^2}}-(x_0-\Delta_{xd})^{1-\frac{2\ln(S)}{\sigma^2}}}{(x_0+\Delta_{xu})^{1-\frac{2\ln(S)}{\sigma^2}}-(x_0-\Delta_{xd})^{1-\frac{2\ln(S)}{\sigma^2}}} \text{ and} \tag{10}$$

$$E(T)=\frac{1}{\frac{\sigma^2}{2}-\ln(S)}\left(\ln\left(\frac{x_0}{x_0-\Delta_{xd}}\right)-\frac{1-\left(\frac{x_0}{x_0-\Delta_{xd}}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}{1-\left(\frac{x_0+\Delta_{xu}}{x_0-\Delta_{xd}}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}\ln\left(\frac{x_0+\Delta_{xu}}{x_0-\Delta_{xd}}\right)\right), \text{ where} \tag{11}$$

$$S=\frac{x_u y_0}{x_0 y_d}.$$

In this model decision boundaries are defined with the help of casualties. The upper and lower decision boundaries for the attacker and the defender are defined as

$$X_u=x_0+d_{xu}(x_0-x_u),\ X_d=x_0-d_{xd}(x_0-x_u), \text{ and}$$

$$Y_u=y_0+d_{yu}(y_0-y_d),\ Y_d=y_0-d_{yd}(y_0-y_d). \tag{12}$$

Decision boundaries are calculated for each battle with their individual force size values. In attacker view we need X_u and X_d and in defender view Y_u and Y_d. The attacker view and the defender view give comparable results but not exactly the same. Because both views must give similar results, different parameter values for individual battles should be used, but these empirical values are not available at combat level. To avoid this difference we take the attacker view when $S>1$ and the defender view when $S<1$. This makes the calculations more consistent and symmetrical with respect to S. The corresponding choice is applied also for expected durations. The knowledge about the winner of a battle is not used. When $S=1$ we can use either the attacker view or the defender view. The model of Equation (12) for decision boundaries gives the probability to win a battle as

$$P=\frac{(x_0)^m-(x_0-d_{xd}(x_0-x_u))^m}{(x_0+d_{xu}(x_0-x_u))^m-(x_0-d_{xd}(x_0-x_u))^m}=$$

$$\frac{\left(\frac{x_0}{x_0+d_{xu}(x_0-x_u)}\right)^{1-\frac{2\ln(S)}{\sigma^2}}\left(1-\left(\frac{x_0-d_{xu}(x_0-x_u)}{x_0}\right)^{1-\frac{2\ln(S)}{\sigma^2}}\right)}{1-\left(\frac{x_0-d_{xu}(x_0-x_u)}{x_0+d_{xu}(x_0-x_u)}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}, \text{ when } S>=1, \text{ and} \tag{13}$$

$$P=1-\frac{(y_0)^{-m}-(y_0-d_{yd}(y_0-y_d))^{-m}}{(y_0+d_{yu}(y_0-y_d))^{-m}-(y_0-d_{yd}(y_0-y_d))^{-m}}=\frac{1-\left(\frac{y_0+d_{yu}(y_0-y_d)}{y_0}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}{1-\left(\frac{y_0+d_{yu}(y_0-y_d)}{y_0-d_{yd}(y_0-y_d)}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}, \text{ when } S<1. \tag{14}$$

For expected duration we have

$$E(T)=\frac{1}{\frac{\sigma^2}{2}-\ln(S)}\left(\ln\left(\frac{x_0}{x_0-d_{xd}(x_0-x_u)}\right)-\frac{1-\left(\frac{x_0}{x_0-d_{xd}(x_0-x_u)}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}{1-\left(\frac{x_0+d_{xu}(x_0-x_u)}{x_0-d_{xd}(x_0-x_u)}\right)^{1-\frac{2\ln(S)}{\sigma^2}}}\ln\left(\frac{x_0+d_{xu}(x_0-x_u)}{x_0-d_{xd}(x_0-x_u)}\right)\right), \tag{15}$$

when $S >= 1$, and

$$E(T) = \frac{1}{\frac{\sigma^2}{2} + \ln(S)} \left(\ln\left(\frac{y_0}{y_0 - d_{yd}(y_0 - y_d)} \right) - \frac{1 - \left(\frac{y_0}{y_0 - d_{yd}(y_0 - y_d)} \right)^{1 + \frac{2\ln(S)}{\sigma^2}}}{1 - \left(\frac{y_0 + d_{yu}(y_0 - y_d)}{y_0 - d_{yd}(y_0 - y_d)} \right)^{1 + \frac{2\ln(S)}{\sigma^2}}} \ln\left(\frac{y_0 + d_{yu}(y_0 - y_d)}{y_0 - d_{yd}(y_0 - y_d)} \right) \right), \quad (16)$$

when $S < 1$

In this approach decision boundaries are constant which is related to the definition of drift term of a geometric Brownian motion describing the combined attrition process of an attacker and a defender.

Comparison with the empirical data

In this section we compare the results of the model presented in the previous section with the empirical combat data (Hartley, 2001). The quality of the empirical data is not high and also the data is scarce for low and high values of $S \notin (0.9, 1.1)$. Also a number of battles have very long duration or very extremely high casualties. As said before, the same durations can be achieved with high variance values or narrow decision boundaries and low variance values or wide decision boundaries. As consequence, it is difficult to fit the parameters of the model with the empirical data. In the following we compare the theoretical results, probabilities and durations, with the empirical data and try to find typical parameter values that are consistent with the theory.

We proceed in three steps. First, we compute the distribution of durations as a function of S and make conclusions which values of variance agree with the empirical data on an aggregate level. On the basis of the first step, we choose one value for the variance and set all the four parameter values for decision boundaries as one. In the third step, we adjust the model parameters, variances and decision boundaries, to find a set of parameters that can describe most of the empirical data. Parameters for individual battles can vary a lot, and in modeling different variances and decision boundaries should be used for each battle. Our goal is to study macroscopic combat effects on an aggregate level. We have two research questions: Is it possible to find a set of model parameter values that describe aggregate empirical results? If yes, can we make conclusions about characteristics of general combat effects? First, we investigate whether distribution of empirical and theoretical duration values of Model 2 agree with each other. In Figure 2 durations for the empirical data points have been calculated from theoretical Equations (15-16) for the three different variance values $\sigma^2 = 0.0025$, $\sigma^2 = 0.01$ and $\sigma^2 = 0.025$. Here, we have set the decision boundary parameters as $d_{xu} = d_{xd} = d_{yu} = d_{yd} = 1.0$. Figure 3 shows the corresponding empirical results (Hartley, 2001) as a function of S. Comparing Figure 2 with Figure 3 an estimation of a typical variance value matching the distributions of aggregate level data can be made.

In Figure 4A the probabilities from Equations (13-14) for attackers to win a battle are shown for the empirical data points with variance $\sigma^2 = 0.0025$ and symmetrical decision boundaries of parameter values $d_{xu} = d_{xd} = d_{yu} = d_{yd} = 1.0$. In Figure 4A the dotted curve shows the probability results calculated with an empirical formula in (Hartley, 2001). To have a check the solid curve shows the same probabilities calculated directly from the empirical data set. These two curves are close to each other which indicate that the empirical data is correctly represented in Figure 4A. The curves also provide rough error estimation for the empirical data.

Figure 4A shows discrepancy of probabilities between the model and the empirical data. With the values of $S \in (0.9, 0.0)$ the model and the data agree at an average level but with values $S \in (0.0, 1.05)$ the model predicts too low values for the probabilities to win a battle. This would indicate that attacker and defender sides are not symmetrical with each other like some earlier studies have indicated (Willard, 1962; Helmbold, 1989; Hartley, 2001).

Figures 4B and 4C show durations as a function of S when $S \in (0.9, 1.1)$ and $S \in (0.5, 2.0)$ respectively. The dotted curves show the empirical observations and the solid curves show the values calculated from Equations (15-16). When $S \in (0.9, 1.1)$ the theoretical curves are above the empirical curves. A higher variance would give somewhat better results but that would make the probabilities in Figure 4A less compatible with the empirical values. Taking into account high uncertainty of the empirical data, the variance value of $\sigma^2 = 0.0025$ can be regarded as a rough compromise between Figures 4A and 4B.

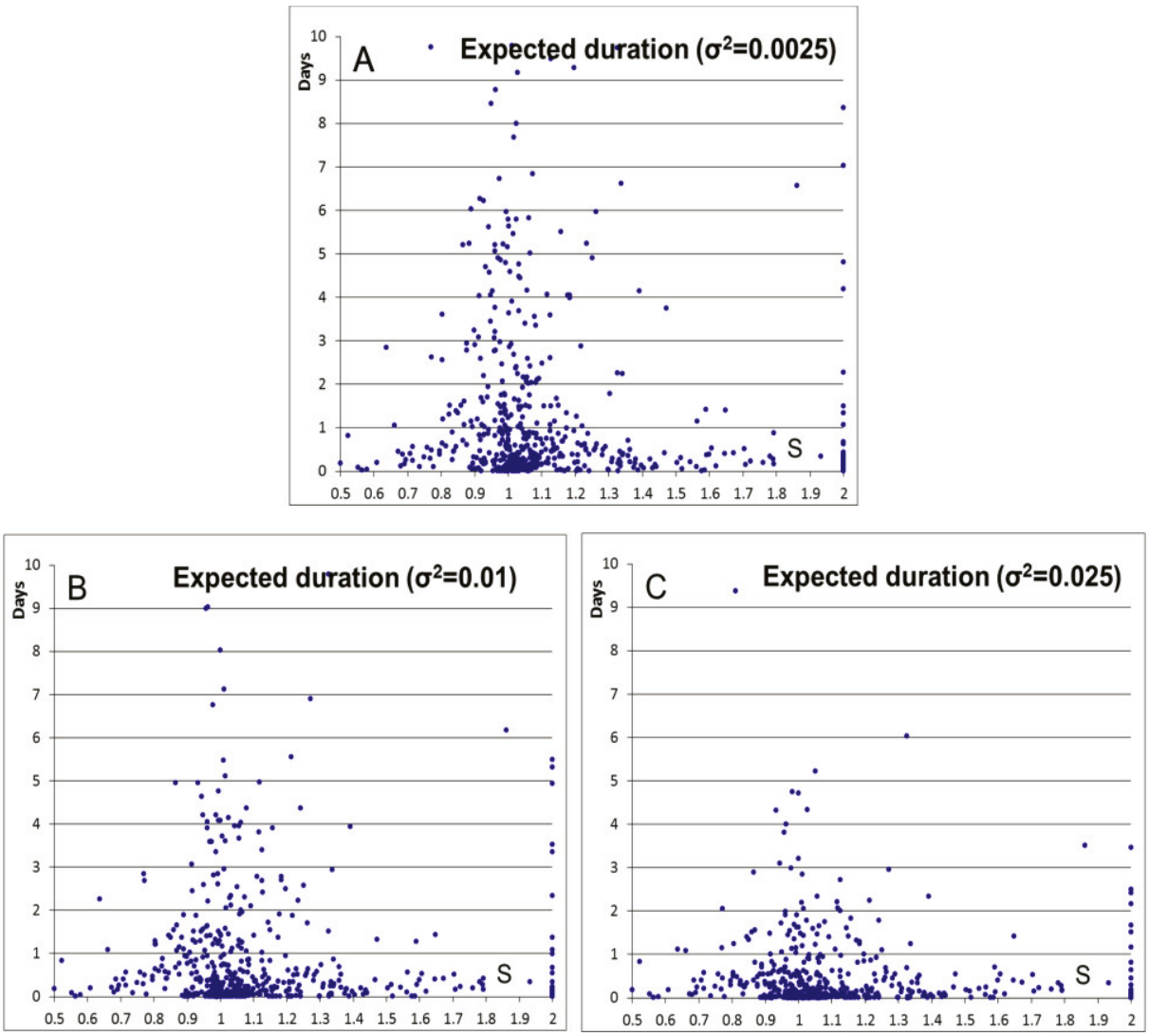

Figure 2. Expected durations from Equations (15-16) with $d_{xu} = d_{xd} = d_{yu} = d_{yd} = 1.0$ and different values of variance σ^2. A) $\sigma^2 = 0.0025$, B) $\sigma^2 = 0.01$, and C) $\sigma^2 = 0.025$.

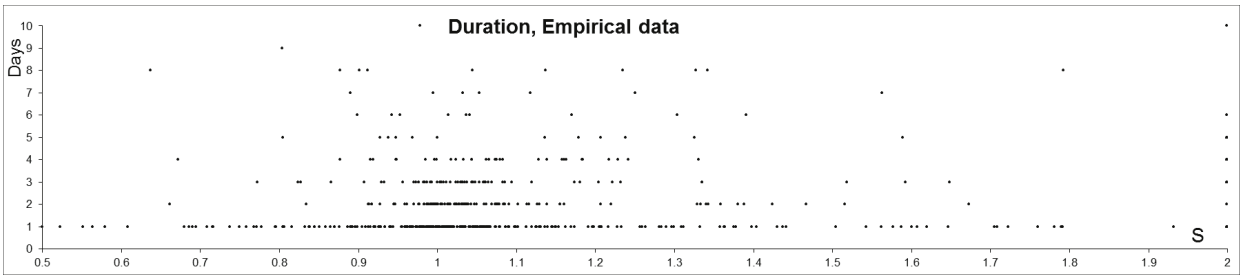

Figure 3. Empirical durations directly from the data set published in (Hartley, 2001) as a function of S.

Figure 4C shows the same quantities as Figure 4B at a wider interval. At higher values of S the theoretical curve is usually below the empirical curve. The interpretation is that when $S >> 1$ the defender is more resilient, in the sense of not surrendering, when compared with battles when $S \approx 1$. The same phenomenon can be observed when $S << 1$, but it is not so apparent, and also less empirical data is available than for high values of S. Lower variance for $S >> 1$ than for $S \approx 1$ is the second possible factor that could explain the discrepancy between the model and the empirical data. We conclude from Figure 1A of Model 1 that different values of variance σ^2 provide almost similar results for low and high values of S. We assume that on an aggregate level, the explaining factors are mostly decision boundaries rather than variances (at a detailed level of individual battles this generalization is not valid).

Figure 4C illustrates that the variance value of $\sigma^2 = 0.0025$ together with the symmetrical decision boundaries parameters $d_{xu} = d_{xd} = d_{yu} = d_{yd} = 1.0$. are not appropriate when $S > 1.05$. The model parameter values in Figure 4 serve as a starting point for determining more appropriate parameter values explaining both probabilities and expected durations as a function of S.

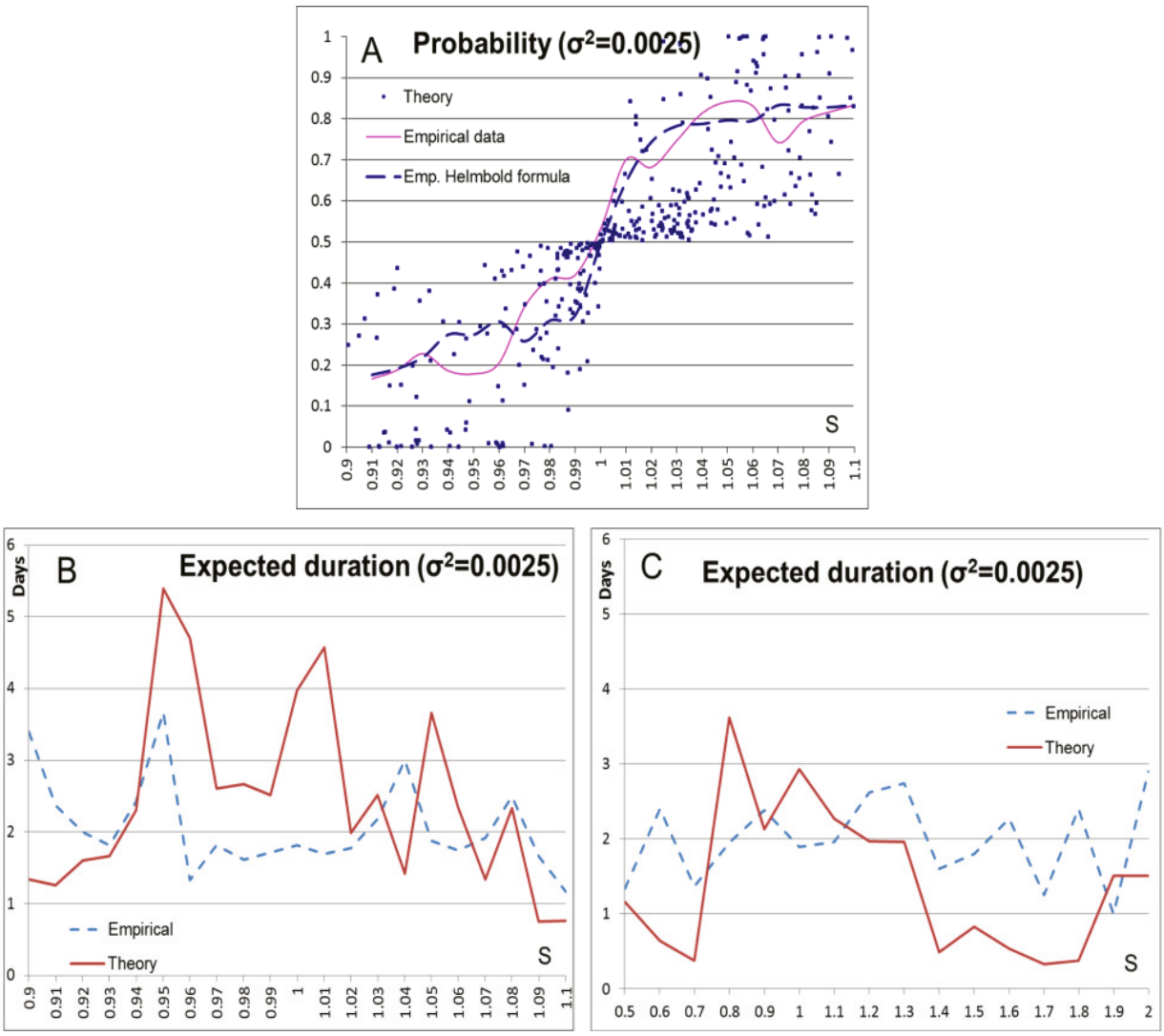

Figure 4. Results for $\sigma^2 = 0.0025$ and symmetrical decision boundary parameter values $d_{xu} = d_{xd} = d_{yu} = d_{yd} = 1.0$. A) Dots show theoretical probabilities to win a battle. Solid curve shows the empirical data from (Hartley, 2001) and dotted curve shows values of an empirical formula in (Hartley, 2001), B) Solid curve shows theoretical expected durations and dotted curve shows the average empirical data, C) As in B for a wider interval of S.

Figures 5 shows the corresponding results with symmetrical narrower decision boundaries $d_{xu} = d_{xd} = d_{yu} = d_{yd} = 0.8$. This makes the duration results in Figure 5B more compliant with the empirical values for $S \in (0.94, 1.06)$. Still, the empirical and theoretical values are different for probabilities in Figure 5A when $S \in (0.0, 1.05)$ and for expected durations in Figure 5C when $S > 1.3$. In other words, distinct intervals of S may exist where different model parameters should be used. A fact should be taken into account when the empirical data set is used: Durations have been documented as integer values. This may result in biased too high values in empirical data.

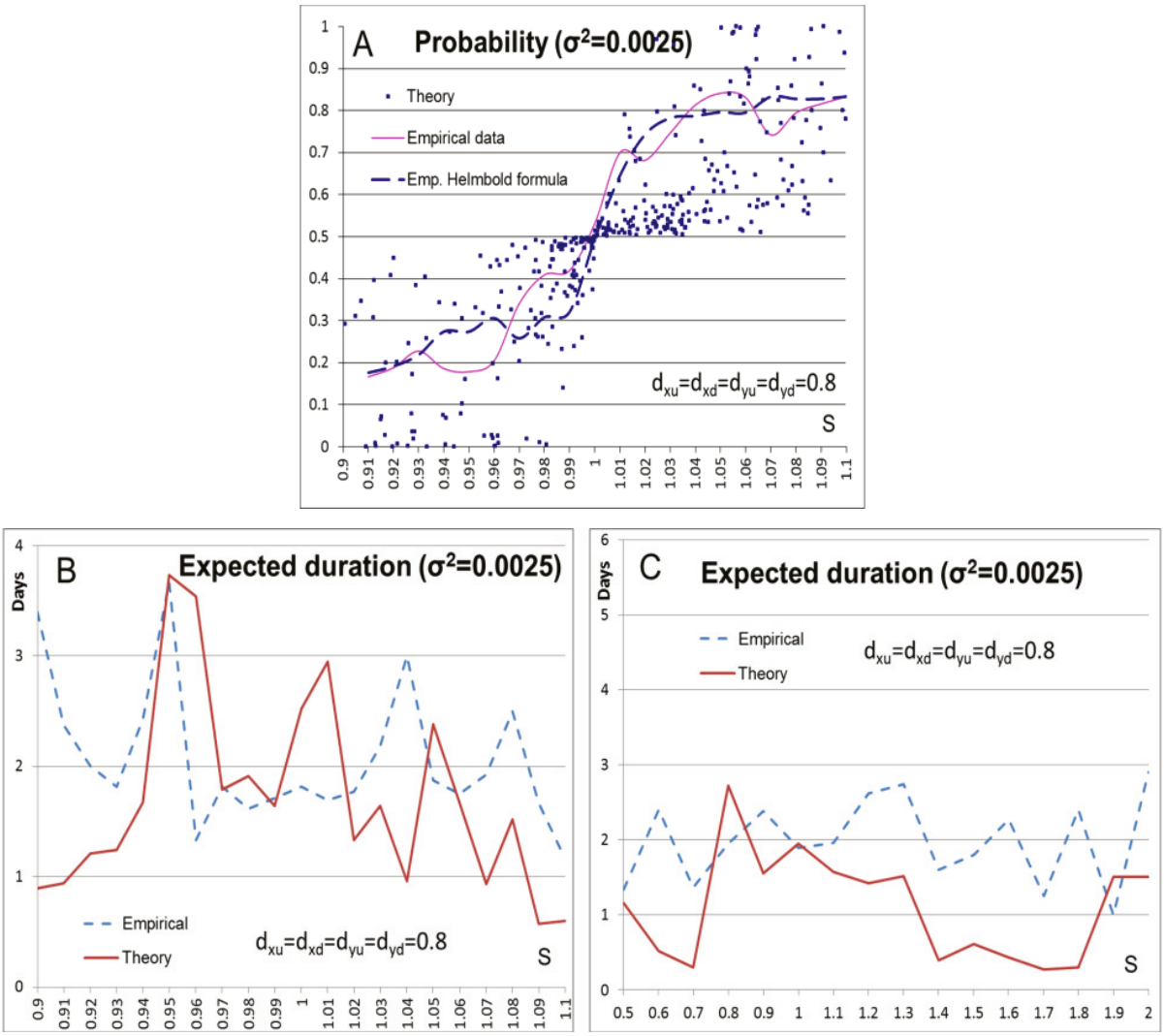

Figure 5. Corresponding results as in Figure 4 for variance $\sigma^2 = 0.0025$ and decision boundaries 20 % closer.

Now we apply the idea of Figure 1C and use asymmetrical decision boundaries to lower the theoretical curve when $S >> 1$. Figures 6A and 6B show the results with the variance value of $\sigma^2 = 0.005$ and with the asymmetrical decision boundary parameters of d_{xu}=1.0, d_{xd}=2.0, d_{yu}=0.8, d_{yd}=1.0 The value of variance is a little higher than in Figure 4. The lower decision boundary values for attackers have been lowered two times lower with respect to the corresponding symmetrical values. Now the empirical and theoretical probability values in Figure 6A are closer to each other in the intermediate interval of $S \in (0.9, 1.1)$. The discontinuity in Figure 6A is irrelevant when we consider S values not very close to one. At the same time, the theoretical expected duration values are also more consistent with the empirical data. However, at low and high values of $S \notin (0.7, 1.4)$ the same discrepancy as in Figure 4 still exists in Figure 6B.

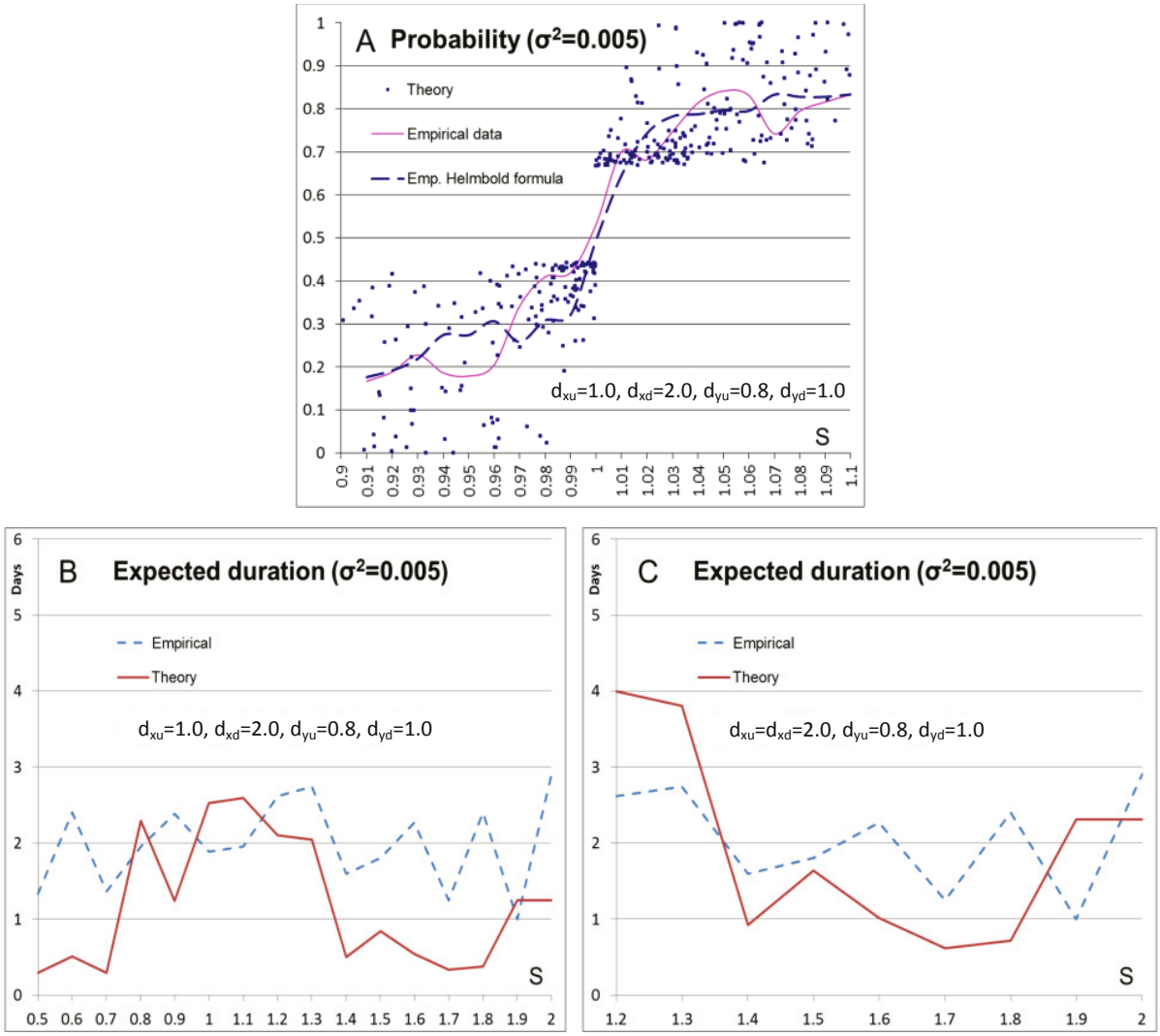

Figure 6. Results with $\sigma^2 = 0.005$ and different decision boundaries. A) Probabilities with asymmetrical decision boundaries d_{xu}=1.0, d_{xd}=2.0, d_{yu}=0.8, d_{yd}=1.0, B) Expected durations with asymmetrical decision boundaries as in A with d_{xu}=1.0, d_{xd}=2.0, d_{yu}=0.8, d_{yd}=1.0, C) Expected durations with decision boundaries d_{xu}=d_{xd}=2.0, d_{yu}=0.8, d_{yd}=1.0 for $S > 1.2$.

Figure 6C shows results with the variance value of $\sigma^2 = 0.005$ and even wider decision boundaries with the parameters of d_{xu}=d_{xd}=2.0, d_{yu}=0.8, d_{yd}=1.0 The parameter values are symmetrical for attacker decision boundaries as there is no direct evidence for asymmetrical decision boundaries for high values of $S > 1.3$. As the empirical data is scarce and diverse for these extreme events, it is not possible to make definite conclusions about possible asymmetry. However, wider decision boundaries can explain longer expected durations.

Summary

We have constructed a stochastic model describing combat effects based on geometric Brownian motion and decision boundaries. Geometric Brownian motion describes force sizes and decision boundaries describe decision rules for ending a battle. The model uses constant decision boundaries. In the terminology of stochastic analysis decision boundaries are called stopping boundaries. The stochastic process of a battle is defined as a difference between the attacker attrition process and the defender attrition process.

Parameters of the model describing a battle are variance and two decision boundaries for both of the opposing sides. The model allows different parameter values for attackers and defenders. Modelling is done at detailed level of individual battles but results are studied on an aggregate level. This is because the available empirical database has no data about variances or decision rules for individual battles. Our goal was to investigate general characteristics of combats by comparing theoretical results and empirical attrition data.

Input variables of the model are the force strengths at the beginning and at the end of a battle. In the model higher variance values and wider decision boundaries are needed when one of the two forces is superior to the other. The interpretation is that on an aggregate level intense fighting, higher variance and resilient defenders are characteristic (or at least some of them) features in majority of these battles. In extreme battles, decision boundaries are more important modelling factors than variances in explaining longer durations of empirical data.

In addition, the model suggests asymmetrical decision boundaries when attackers are superior to defenders. This indicates that attackers have an extra advantage with respect to defenders. Asymmetrical features between attackers and defenders are discovered by comparing empirical data with the modelling results of probabilities and durations.

In this paper, analytical closed form formulas are presented for calculating the probability of victory and the value of expected duration of a battle. The probability to win a battle or a war is a combination of numerous different decisions on many levels of decision making. The modelling methods can be used for developing more detailed sub-models for decision rules. These techniques can have applications in decision making situations in other contexts than military.

References

Dupuy, T. N. (1985). Numbers, Predictions & War. Hero books.

Hartley III, D. S. (2001). Topics in Operations Research: Predicting Combat Effects. Institute for Operations Research and the Management Sciences.

Helmbold, R. L. (1989). Do Battles and Wars Have a Common Relationship Between Casualties and Victory. Presented at the 57th MORS Symposium, Forth Leavenworth, Kansas.

Kim, D., Moon, H., Park, D., Shin, H. (2017). An efficient approximate solution for stochastic Lanchester models. Journal of the Operational Research Society 68(11), 1470-1481.

Kuikka, V. (2015). A Combat Equation Derived from Stochastic Modeling of Attrition Data, Military Operations Research, Vol. 20, No. 3, 49-69.

Kuikka, V., (2016). Number of System Units Optimizing the Capability Requirements through Multiple System Capabilities, Journal of Applied Operational Research, Vol. 8, No. 1, 26–41.

Lanchester, F. W. (1914). Aircraft in Warfare: The Dawn of the Forth Arm – No. V., The Principle of Concentration, Engineering, Vol. 98, 422–423.

Lappi, E., Pakkanen, M., Åkesson, B. (2012). An approximative method of simulating a duel. In Proceedings of the Winter Simulation Conference, WSC'12, pp. 1–10.

Perry, N. (2011). Applications of Historical Analyses in Combat Modelling. Technical Report DSTO-TR-2643.

Taylor, H. M., M. Karlin, (1998). An Introduction to Stochastic Modeling 3rd Edition, Academic Press.

Willard, D. (1962). Lanchester as Force in History: An Analysis of Land Battles of the Years 1618-1905. No AD297375, DTIC, Alexandria, Virginia.

Willmott, P. (2000). Quantitative Finance, Vol. 1, p. 163. John Wiley & Sons.

VI

Metrics for Networked Systems Design in a Network-Centric Warfare Context

Vesa Kuikka

Journal of Battlefield Technology 10(3), 2007, pp. 27-36

METRICS FOR NETWORKED SYSTEMS DESIGN IN A NETWORK-CENTRIC WARFARE CONTEXT

Vesa Kuikka[1]

Abstract. This paper presents modelling techniques for networked systems. A network metric is suggested for evaluation and comparison of different networked systems in a network-centric warfare context with random or targeted attacks against the network structure. A method is also presented to calculate the threshold values of link failure probabilities where it is optimal for an attacker to shift targeted attacks against links of lower degree nodes. This information is vital for the defender in planning and constructing more robust networks against targeted attacks.

1. INTRODUCTION

This paper discusses modelling networked systems of services and users. Expressions for robustness are derived for realistic military situations. The methods are not limited to low failure probabilities so they are suitable for vulnerable and robust systems.

The goal of this paper is to develop a general framework for studying and comparing different networked systems when network links (edges) and nodes (vertices) become non-functional. The models presented here have a common viewpoint that one connection is enough for two parties to communicate, two connections are enough for three parties and so on. Here we use the term connection when there is a path between nodes—that is, connectivity is maintained. When all the nodes break down during the same time period, the system is down for the next time period. It is possible that during the repair period the nodes are destroyed again.

In the basic model, all the functional nodes in the system can do the same job. This is crucial when network metrics are calculated and makes a big difference to another definition of network metrics used in [6], where all the possible connections and nodes have been summed and weighted with values of links and nodes.

A networked system comprises three different elements: network links, functional network nodes, and technical network nodes. Functional network nodes are computer centres, command centres, users of networked services, and so on. Technical nodes are, for example, network routers and there are no applications for end users. Destroying a functional network node is usually more serious than destroying a technical network node. [7]

The framework consists of the following four steps:

1) Evaluate the failure probability p for a general target, a node or a link.
2) Evaluate the failure probability for a system of several nodes (Section 3).
3) Evaluate the failure probability for a network connection (several links) (Sections 4 and 5).
4) Evaluate the failure probability for a combined system of nodes and connections (Section 6).

In the first step the failure probability is evaluated for one target where firing theory is the tool if weapons are used against infrastructure [15]. Computer viruses and technical errors have to be taken into account by different methods. At the end of this step failure probabilities for functional nodes and links have been evaluated.

In step two, scenarios for one or several nodes are studied where, in the case of a node breakdown, a task transfer to another node can occur. The receiving node can be a standby node or a non-standby node. Task transfer to a standby node happens automatically and immediately after the failure. Task transfer to a non-standby node can occur after manual operations or rebuilding of the system. The main difference is that a transition to a standby node happens directly without delays and a transition to a non-standby node requires a preparation phase.

The goal in Section 3 is to construct a model sufficiently general to allow different modifications. The theory behind is discrete or continuous time Markov processes [5]. For Markov processes future states of the system are not influenced by the past of the system, only the current state is important. The limiting distribution is calculated as a first model but given the initial state, time dependence can be studied easily. Markov matrices serve also as a visualization tool for understanding the system's behaviour.

Step three is analogous to step two but the system is composed of network links. The probability for a connection between functional nodes is evaluated and a method is presented for the computation. Section 4 introduces some basic concepts of network connectivity. Section 5 presents a method to compute a measure of robustness for military networks or civilian networks under the threat of terrorist attacks. In the model attacks are not random but targeted so as to cause maximum damage to a network and unequal failure probabilities for network links are considered. Markov matrices or limiting processes are not used in Section 5 but the results can be utilized in Section 6 for limiting distributions. Alternatively, time-dependent solutions can be calculated by iterating waves of attacks with the methods of Section 5 alone.

In step four the system is regarded as a collection of functional nodes and links. The total failure probability for the system can be calculated. This is also a measure for comparing networks and networked systems. The survival probability of the system or the entire network is suggested as a new network metric (Equation (6.4)).

In the long run, if no repair takes place, the system will eventually break down. But if the number of spare parts or nodes is sufficiently large, compared to the time of hostile activity, the networked system can still operate until the end of the war. In this case the probability is fairly easy to compute. The probability distribution for the system in operation is given by the binomial distribution.

[1] Finnish Defence Forces, CIS Centre Development Division, Kutojankulma 2, PL 210, 02631 Espoo, Finland.

ISSN 1440-5113

Models 1 and 2, in Section 3, give the limiting distributions π_0 and π for the probabilities for the system not in operation and in operation respectively. In the models Q is the ratio of uptime to downtime of the system and it is a function of the number of replaceable nodes of the system. If in the system planning phase, a requirement for Q is given, a lower bound for the number of nodes can be solved.

2. RELATED WORK

Network robustness has been studied in the areas of graph theory, telecommunications network reliability, and percolation theory. Recently there has been some interest in combining networking and information systems in modelling military command systems. References to related work are given in this section and comparisons to this paper are provided in Section 7.

In [1,2] robustness of network topologies are examined with graph-theoretic concepts. Robustness of the topology will come from the presence of alternate paths, which ensure that communication remains possible in spite of damage to the network. The authors suggest that it is important for robust networks to satisfy two conditions: node-similarity (or symmetry) and optimal connectivity. Optimal connectivity is defined by $\kappa = \lambda = d_{min}$, where κ is node connectivity, λ is link connectivity, and d_{min} is the minimum degree of the graph. Node (link) connectivity, κ, is the smallest number of nodes (links) whose removal results in a disconnected or single-node graph. The degree of a node is the number of outgoing links and d_{min} is the minimum degree of the graph. Node connectivity is suggested as a measure for network robustness.

Modelling random failures have been studied, for example, in References [9,10,12]. In [12] independent link failure and repair probabilities are incorporated into network reliability calculations. Using Markov chains is common in reliability analysis and the method has been used in [12]. Different targeted attack strategies against networks have been examined in [4] where the importance of the network structure changes during targeted attacks is emphasized. Betweenness centrality (defined in Equations 4 and 5 in [4]) and degree of the node (link) have been used as the selection criteria for targeted attacks against nodes (links). Betweenness and degree are correlated but not always removing nodes (links) with high centrality or high degree have the same effect on the network robustness.

There has been much interest in random and targeted attacks against large networks in percolation theory [13]. In targeted attacks nodes are removed in the order of descending degree and in random attacks nodes are removed in random. A typical question is whether or not there exists a 'giant component' in the limit of large system size. The size of the largest connected subgraph is an important quantity and it has been used also in [4] as a measure of network robustness. Conditions for the critical fraction of nodes p_c that needs to be removed before the network disintegrates have been derived for various networks structures such as random and power law networks. Optimization of network robustness to waves of targeted (fraction p_t) and random (fraction p_r) attacks gives a bimodal degree distribution [13]. The fraction r of the nodes has degree $k_2 = (\langle k \rangle - 1 + r)/r$ and the remainder of the nodes have degree $k_1 = 1$, where $\langle k \rangle$ is the average degree of all the nodes and the optimal value of r is of the order of p_t / p_t for $p_t / p_r << 1$.

In [6] a connectivity measure of a military network is defined as the sum of the value of all the nodes and their connections scaled by the lengths of the routes and their directionality. Metrics for assessing the value of networking and information in the battlefield have been discussed with a method where network metrics and decision-action rate have been related. In [7,8,11] the measure of total network performance quantifies the effects of information sharing across a cluster on information completeness and accuracy, whereas plecticity measures positive and negative effects of redundant information and the degree of information access.

For networks of more than about 25 links optimized algorithms or simulation methods are necessary because of time and memory complexity. Examples of algorithms are given in [3,14].

3. RECOVERY OF COMMAND CENTRES

Because military systems may be targets for hostile attacks special actions are necessary to increase robustness of systems. One obvious way is to replicate application servers. If one copy of the system is destroyed another takes the task and so on. In this section we study robustness of the replicated system and in later sections networking is included in the model.

Next we present a mathematical model based on Markov chains for a system of nodes which are destroyed or where a technical error occurs at random times. The actions to get the system operational again are to repair or to replace the node. There are many scenarios that can be described with this model.

Our model is based on Markov chains [5]. The limiting distribution is:

$$\pi_j = \sum_{k=0,N} \pi_k P_{kj}, j = 0,1,...,N \tag{3.1}$$

The distribution is unique for regular transition probability matrices (P_{kj}) [5, Theorem IV.1.1]. The number of states is N+1. The convergence means that, in the long run, the probability of finding the Markov chain in state j, is approximately π_j, no matter in which state the chain began at time 0. The system has reached the stationary state.

Imagine a situation that one command centre is completely destroyed but with the help of advanced technology another command centre takes over in almost real time and users of the system are completely unaware of what has happened. In addition, it is assumed that the nodes not in operation are repaired in the background during the same time period. This is why the number of nodes n is kept constant. In state 0 nodes are still vulnerable and can be destroyed again during the repair period. This assumption may be unrealistic in practice and a more accurate model could be constructed. Later in this paper more refined Models 2 and 2B will be presented where more states of the system will be introduced.

The parameters of the model are p, p_r and p_t where p is the failure probability during the specified time interval, p_r is the conditional probability to repair or replace the system during

the time period and p_t is the conditional probability of task transfer of a node.

The model is a simplification of the real world system in several respects. We assume that the process is a Markov process with constant parameters p, p_t, and p_r. The model for repairing the system is approximate because the distribution for the duration of the repair time is usually not known. An effect of discrete time is that the system always stays one time step or period in the state where it arrives, before it makes another transition

Fast task transfer to another node is studied in Model 1. States 1 and 2 are defined as a combination of several nodes: one or more of the n nodes are not destroyed. The only difference between states 1 and 2 is that they have different history. Transition to state 0 occurs when the last of the n nodes is destroyed. The Markov matrix becomes:

$$P = \begin{vmatrix} 1-(1-p)p_r-\left(1-p^{n-1}\right)p_t & (1-p)p_r & \left(1-p^{n-1}\right)p_t \\ p^n & 1-p^n & 0 \\ p^n & 0 & 1-p^n \end{vmatrix} \quad (3.2)$$

The ratio of uptime to downtime of the system $Q(n)$ can be solved from the limiting distribution of (3.1)

$$Q(n) = \frac{\pi}{\pi_0} = \frac{(1-p)p_r + \left(1-p^{n-1}\right)p_t}{p^n} \quad (3.3)$$

The variables π_0 and $\pi = \pi_1 + \pi_2 = 1 - \pi_0$ give the probabilities for the system not operating and operating:

$$\pi_0 = \frac{1}{1+Q(n)} \text{ and } \pi = \frac{Q(n)}{1+Q(n)}.$$

The values with parameters $p_r = p_t = 0.01$ of $Q(n)$, n=2, 3, 4, 5 have been shown as a function of the failure probability p in Figure 1.

In [8] a network penalty factor has been introduced to account for negative effects of information overload. Heuristic logistic shape equation $g(C)$ with parameters a and b has been introduced to account for additional network connectivity:

$$g(C) = \left(e^{a+bC}\right) / \left(1+e^{a+bC}\right)$$

C is the total number of network links accessed by nodes on the critical path. The approach is opposite to this presentation where vulnerability is considered as a central phenomenon in networked systems. In our model no heuristic penalty functions are necessary because additional nodes and links are redundant and the only effect is to improve reliability of the network. (3.3) can be compared to the logistic penalty function of [8]. (3.4) gives n as a function of $Q(n)$

$$n = \frac{\ln\left(\dfrac{p_r + p_t - pp_r}{pQ + p_t}\right)}{\ln p} + 1 \quad (3.4)$$

For an example $Q=50$ (2% not in operation) and p=0.2, p_r=p_t=0.1, the result from (3.4) is $n \approx 3.5$—that is, four nodes are needed to maintain the system at 98% capacity. For a low value of $Q=5$ (20% not in operation) two nodes are almost enough to maintain the system at 80% level. In Figure 1 these examples have been marked with dots.

In Model 2, the state space is larger than before, also the limiting probabilities for n, n–1, n–2, …, 1 nodes in operation will be calculated. In the basic model, constant repair time is assumed. For example, if there are three nodes in the system the Markov matrix is:

$$P = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & p & 1-p \\ 0 & p^2 & 2p(1-p) & (1-p)^2 \\ p^3 & 3p^2(1-p) & 3p(1-p)^2 & (1-p)^3 \end{vmatrix} \quad (3.5)$$

The elements of the matrix are the probabilities $P\{X=k\} = \binom{n}{k}(1-p)^k p^{(n-k)}$ of binomial distribution with parameters n and p. The states of the system are labelled according to the number of nodes in operation i=0, 1, 2, 3. The limiting distribution is:

$$\pi_0 = \frac{p^3}{(1+p)^3}, \pi_1 = \frac{3p^2}{(1+p)^3}, \pi_2 = \frac{3p}{(1+p)^3}, \pi_3 = \frac{1}{(1+p)^3}$$

The general equations for n nodes for the limiting distribution are:

$$\pi_k = \sum_{i=n-k,n} \frac{i!}{(n-k)!(i-n+k)!} p^{n-k}(1-p)^{i-n+k}\pi_i \quad (3.6)$$

and the normalization equation is $\sum_{i=0,n} \pi_i = 1$. By Theorem [5, IV.1.1] the solution is unique for $0 < p < 1$ and we get:

$$\pi_i = \binom{n}{i} \frac{p^{n-i}}{(1+p)^n}, i = 0,...n \quad (3.7)$$

where: $\binom{n}{i} = \dfrac{n!}{(n-i)!\,i!}$

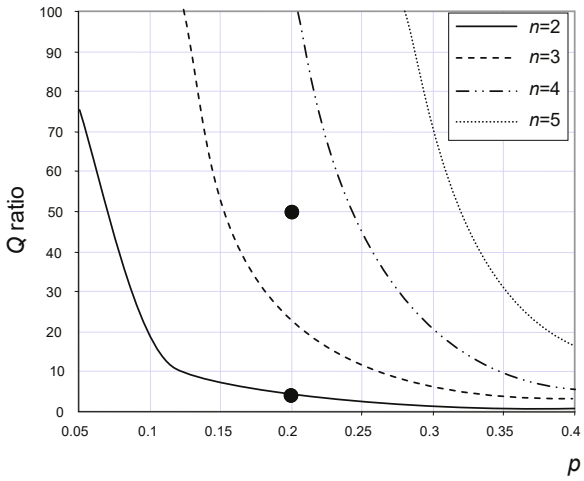

Figure 1. (Model 1, parameters $p_r = p_t = 0.1$). Ratio Q as a function of the failure probability p for one node, with 2, 3, 4 and 5 nodes in the system. The examples given in the text have been marked with dots.

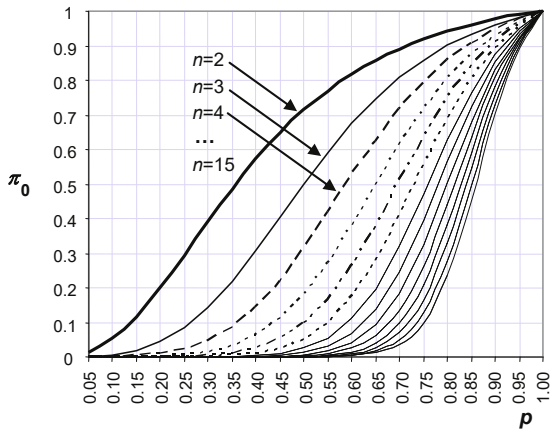

Figure 2a. (Model 1, parameters p_r=p_t=0.1). Limiting failure probability for the system as a function of the failure probability *p* for one node and the number of nodes *n*=1,…,15.

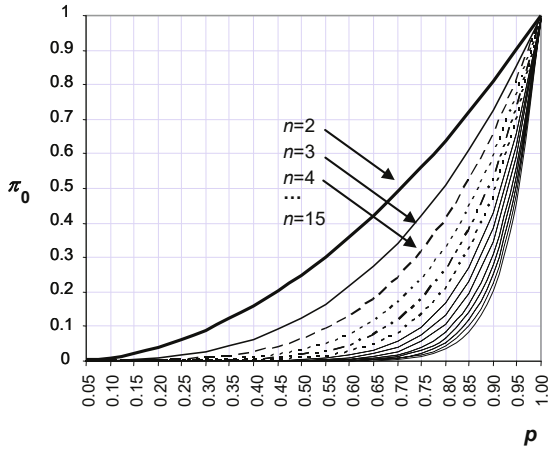

Figure 2b. (Model 2, parameters p_r=p_t=0.1). Limiting failure probability for the system as a function of the failure probability *p* for one node and the number of nodes *n*=1,...15.

Equation (3.7) can be verified by substitution into (3.6). The factor $1/(1+p)^n$ is for normalization. When $p<1$ and $n \to \infty$ the probability of no operation $\pi_0 \to 0$ for Models 1 and 2. The striking difference is that $\pi_0 \to 1$, in Model 1, and $\pi_0 \to 1/2^n$, in Model 2, when $p \to 1$. As the number of nodes increases the probability for a system not in operation goes to zero in Model 2. The difference is due to the higher repair activity of Model 2 compared to Model 1. In Figures (2a) and (2b) Models 1 and 2 are compared for the parameter values $p_r = p_t = 0.1$. The probabilities of no operation have been shown as a function of the failure probability p with n=1,..,15 replaceable nodes in the system.

For realistic applications more accurate models may be necessary. In Model 2B, the repair time of nodes will be taken into account. The 4×4-matrix corresponding to (3.5) is:

$$P = \begin{vmatrix} (1-p_r)^3 & 3p_r(1-p_r)^2 & \dots & \dots & \dots & 3p_r^2(1-p_r) & p_r^3 \\ (1-p_r)^2 p & (1-p_r)^2(1-p)+2p_r(1-p_r)p & \dots & \dots & \dots & 2(1-p_r)p_r(1-p)+p_r^2 p & p_r^2(1-p) \\ (1-p_r)p^2 & 2(1-p_r)p(1-p)+p_r p^2 & \dots & \dots & \dots & (1-p_r)(1-p)^2+2p_r p(1-p) & p_r(1-p)^2 \\ p^3 & 3p^2(1-p) & \dots & \dots & \dots & 3p(1-p)^2 & (1-p)^3 \end{vmatrix} \quad (3.8)$$

The limiting distribution for n nodes in the system is:

$$\pi_i = \binom{n}{i} \frac{p_r^{\,i} p^{n-i}}{(p_r+p)^n}, i = 0,\dots,n \quad (3.9)$$

The theory can be generalized in many ways. For example combined states of nodes and links can be considered or history of the system can be included in the model. Also different failure or repair rates can be incorporated for different number of broken nodes resulting in different values of p and p_r on matrix lines in (3.8).

4. PROBABILITY FOR AN OPERATIVE CONNECTION

Our next step is to connect the command centres in a wide area network. The same model can be applied for connecting computers in a local area network or people with telephone lines and so on. We assume that there is a connection between two sites if there is at least one connection, or path, in operation. Figure 3a shows two simple networks which we take as examples. A general method to calculate the failure probability will be presented based on the first example. [10] In Section 5 the second test network is studied in more detail.

We assume that the failure probability for every link in the graph is p_l during the specified time period and the probability is independent of the length of the link. We investigate the first test network of Figure 3a. To make the situation practical we can list all the combinations of broken links. There are 32 all together, when the case with no broken links, is also counted. The probability that one link is broken is $p_l(1-p_l)^4$, the probability that two links are broken is $p_l^2(1-p_l)^3$ and finally the probability that all the links are broken is p_l^5. There are one, five, ten, ten, five and one combinations for zero, one, two, three, four and five broken links. This is the familiar Pascal triangle rule. The normalization is:

$$\sum_{s=0,5} \binom{5}{s} p_l^{\,s} (1-p_l)^{5-s} = 1$$

In our example if only one link fails there is a connection between all the nodes. For two link failures two there are two possibilities of no connection, and so on. The link connectivity is two for the network. We calculate the

probability of no connection between the marked nodes one and four of the first test network of Figure 3a as:

$$p_L = 1 - \Big[(1-p_l)^5 + 5p_l(1-p_l)^4 + 10p_l^2(1-p_l)^3 + 10p_l^3(1-p_l)^2 + 5p_l^4(1-p_l) + p_l^5 - 2p_l^2(1-p_l)^3 - 8p_l^3(1-p_l)^2 - 5p_l^4(1-p_l) - p_l^5 \Big] = 2p_l^2 + 2p_l^3 - 5p_l^4 + 2p_l^5 \quad (4.1)$$

where, in the parenthesis, the probabilities of no connection have been subtracted. If we think that the failure probability for one link is small, it is a good approximation, to take only the first powers of p_l. In the example $p_L \approx 2p_l^2$ if p_l is small. Generally, for low values of p_l we have $p_L \approx A_C p_l^C$ where C is the link connectivity and A_C is given later in (5.3). In Figure 3b the probability of no connection between nodes one and four for the first test network and nodes four and nine for the second test network are shown as a function of the link failure probability. The connection in the first network is more robust than the connection in the second network.

Note that the theory presented in Section 3 applies also for links in the network. For example from (3.9) the probability of operation is:

$$\pi = 1 - \frac{p^n}{(p_r + p)^n} \approx 1 - \frac{2^n p_l^{2n}}{\left(p_r + 2p_l^2\right)^n} \quad (4.2)$$

and the ratio:

$$Q = \frac{\pi}{\pi_0} = \left(\frac{p_r}{p_L} + 1\right)^n - 1 = \left(\frac{p_r}{2p_l^2 + 2p_l^3 - 5p_l^4 + 2p_l^5} + 1\right)^n - 1 \approx \left(\frac{p_r}{2p_l^2} + 1\right)^n - 1 \quad (4.3)$$

In system design, the lower bound for links n, can be solved from (4.3) when the value of required Q is given.

The probabilities $p_{L,uj}$ should be computed for different connections between nodes u and j in the network. Usually, in real networks, the probabilities $p_{L,uj}$ are different for each connection, as a result of the network structure. This problem is studied for random and not random failure probabilities in Section 5. As an example, consider a system of a user u and two nodes 1 and 2 where the user needs a connection L_{u1} or L_{u2} to one of the nodes. Equation (3.8) for three states transforms into:

$$P = \begin{vmatrix} (1-p_r)^2 & 2p_r(1-p_r) & p_r^2 \\ (1-p_r)p_{L,u1} & p_r p_{L,u1} + (1-p_r)(1-p_{L,u1}) & p_r(1-p_{L,u1}) \\ p_{L,u2}^2 & 2p_{L,u2}(1-p_{L,u2}) & (1-p_{L,u2})^2 \end{vmatrix} \quad (4.4)$$

where $p_{L,u1}$ and $p_{L,u2}$ are the failure probabilities for the connections L_{u1} and L_{u2}. The analysis for the model proceeds as before and the quantities $\pi_0, \pi_1, \pi_2, \pi, Q, \ldots$ can be solved.

5. TARGETED ATTACKS AGAINST NETWORK LINKS

In Section 4 the failure probability p_l is equal for every link of the network. [9,10] This may be a sufficient model if hostile attacks against links are random or every link is attacked with uncertain success. If a network is targeted so as to cause maximum damage we need to go further and construct a model where failure probabilities are unequal for network links. [12]

As a practical application of the theory we show how to calculate the probability of no connection between two nodes, for example, a user and a service in the network. Because we don't expect hostile forces to have exact information about the locations of services we define an average metric for a general network topology. The model works also for specific situations where the locations of users and services are known but the average metric is suitable to show the general behaviour of the network under targeted attacks.

First we construct the mathematical theory. We start from the general equation for the probability $P_{L,ab}$ of no connection between two nodes a and b of a network [10,12]:

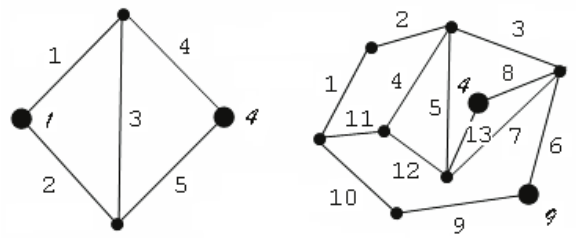

Figure 3a. Two example networks studied in Sections 4 and 5. Links have been labelled. The probabilities of no connection between the marked nodes have been shown in Figure 3b.

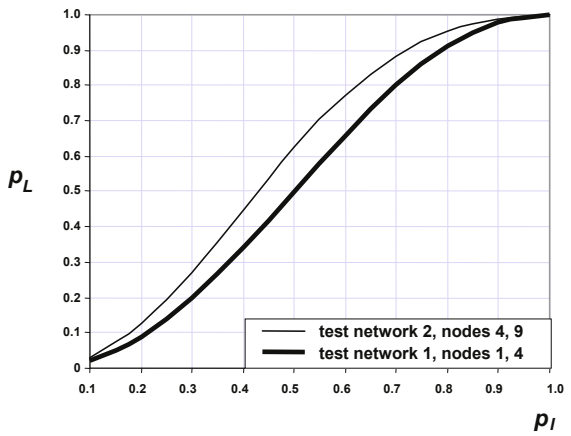

Figure 3b. The probability of no connection between nodes 1 and 4 for the first test network as a function of the link failure probability (thick line). Similarly, the probability of no connection between nodes 4 and 9 for the second test network (thin line) is shown.

$$P_{L,ab} = 1 - [\prod_{i}^{N_L}(1-p_i)I_{ab,\{\}} + \sum_{j}^{N_L}\prod_{\substack{i=1\\ i\neq j}}^{N_L}(1-p_i)p_j I_{ab,\{j\}} + \sum_{\substack{j,m\\ j<m}}^{N_L}\prod_{\substack{k=1\\ k\neq j,m}}^{N_L}(1-p_k)p_j p_m I_{ab,\{j,m\}} + \ldots + \sum_{j}^{N_L}\prod_{\substack{i=1\\ i\neq j}}^{N_L} p_i(1-p_j)I_{ab,\{1,\ldots,j-1,j+1,\ldots,N_L\}}\ ] \tag{5.1}$$

where N_L is the number of links in the network and p_i is the failure probability for link i. The terms in the parenthesis correspond to events of no links destroyed, one link destroyed and so on. The value of the indicator function $I_{ab,\{i,j,\ldots\}}$ is one if there is a connection between nodes a and b when links $i,j,\ldots$ have been removed, otherwise $I_{ab,\{i,j,\ldots\}}$ is zero. Equation (5.1) applies for three node systems if the indicator functions are replaced by corresponding three node terms $I_{abc,\{i,j,\ldots\}}$. If the network is connected or there is a connection between a and b the first indicator function $I_{ab,\{\}}$ is one.

From (5.1) we get, as a special case, when the failure probabilities are equal $p_i=p_l$, $i=1,\ldots,N_L$.

$$P_{L,ab} = \sum_{h=1}^{N_L} A_h p_l^{\,h} \tag{5.2}$$

where:

$$A_h = \sum_{s=0}^{h}(-1)^{h-s}\binom{N_L - s}{N_L - h}H_s \tag{5.3}$$

and H_s is the sum of indicator functions when s links are removed. Equation (5.2) is a polynomial of maximum order of N_L. For example, we get (4.1) with $N_L=5$ and $H_0=1$, $H_1=5$, $H_2=8$, $H_3=2$, $H_4=0$, $H_5=0$. But if the probabilities are not equal (5.1) does not simplify and we have to use it directly.

Before we proceed into results we define a measure for the network robustness as an average over every two node pairs in the network:

$$P_L = \frac{1}{N_L(N_L-1)}\sum_{\substack{i,j=1\\ i\neq j}}^{N_L} P_{L,ij} \tag{5.4}$$

Again, the average P_L is a polynomial in p_i, $i=1,..,N_L$ but for unequal probabilities we compute the numerical values from (5.1).

Our goal is to find for military networks, or civilian networks under terrorist attacks, the strategy for maximum damage. Network topology, link failure probabilities and number of attacks are used to calculate the exact probabilities $P_{L,ab}$ for connections between two nodes a and b of the network. We use the average metrics of (5.4). Usually military networks are small with at most tens of nodes and tens of links. For larger networks (5.1) and (5.2) can be truncated when the wanted accuracy is gained. The first terms are suitable for approximations if the failure probabilities are small. If the failure probabilities are high the last terms of (5.1) are suitable for computation.

In this section, one hostile action against a link is called an attack and attacks are assumed to be independent with probabilities of success p_i against link i. For example, an attack may be a bomb dropped from an airplane or an electromagnetic disturbance against a wireless network link. In our examples we assume that for the first attack the probabilities are the same $p_i = p_l$, $i=1,..,N_L$ and for the k'th attack.

$$P(k) = p_l + P(k-1) - p_l P(k-1) \tag{5.5}$$

where $P(1) = p_l$. If previous attacks affect the success of future attacks the equation for $P(k)$ should be modified correspondingly. The general method, given in this section, still applies without change. Note that for small p_l the 'penalty' term $p_l P(k-1)$ is negligible and for p_l close to one the effect is large. This has a considerable effect on the results. In fact, the network topology and the model for $P(k)$, together, explain the phenomena of the results given below. We denote the number of attacks against the link i of the network as n_i and define the strategy of the attack as the vector $(n_1, n_2, \ldots, n_{N_L})$.

In Figure 4a the average metric of (5.4) for the first network topology of Figure 3a is shown for the four strategies (1,1,1,1,1), (3,2,0,0,0), (1,1,0,1,1), and (2,2,0,0,0). One attack against each link of the network is conducted in the first case, three attacks against the first and two against the second link in the second case and so on. In the last two examples total number of attacks is less than the number of links of the network. In cases where one or more links remain without attacks the values of P_L are less than one when p_l=1—that is, total destruction is not achieved. Naturally, the number of attacks may be larger than the number of links but the most significant results are found from our simple examples. When fever attacks are conducted the attacks must be more effective to reach the same P_L and this is demonstrated with the shift of the last two curves to the right.

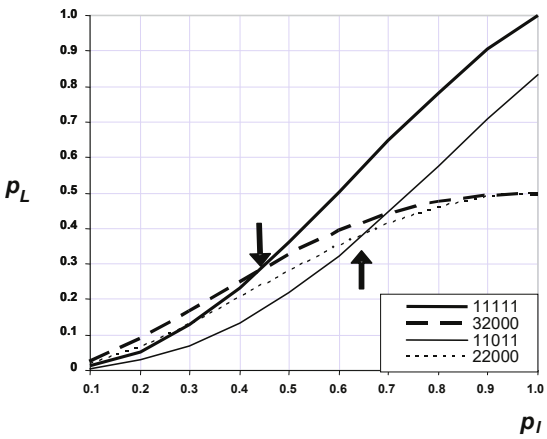

Figure 4a. The average metric of (5.4) for the first topology of Figure 3a is shown for the four strategies. The first two strategies have 5 attacks and the last two have 4 attacks against the first test network. The change of the optimal strategy for the attacker is indicated with arrows for both cases.

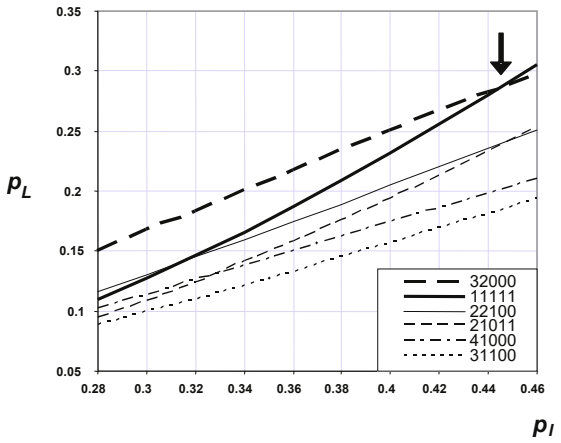

Figure 4b. The average metric of (5.4) for the first topology of Figure 3a is shown for the six strategies when p_I is between 0.28 and 0.46. All the strategies have 5 attacks against the network.

The most important result is that the optimal strategy changes when the probability of the success of the attack changes. The thresholds for five attacks and for four attacks are indicated by the two arrows in the figure. In the case of five attacks the optimal strategy is to attack all the five links if the failure probability is more than 0.45 for one link. If the probability is less than 0.45 more damage is caused if the attacks are shifted to the first and second links or because of symmetry to the links four and five. Exactly the same behaviour is seen from the four attack strategies where the optimal strategy is one attack for the links one, two, four and five—that is, (1,1,0,1,1) when p_I is larger than 0.64. Note, that the strategy (1,1,1,1,0) is not optimal for any value of p_I because nodes one and four are of degree two and nodes two and three of degree three. The priority is on the weakest links. We still find something interesting, for values of p_I less than 0.3 the four attack strategy (2,2,0,0,0) is a little more effective than the five attack strategy (1,1,1,1,1). This is due to the interplay between the network topology and the model for $P(k)$ in (5.5).

In Figure 4b some relevant five attack strategies are shown for p_I between 0.28 and 0.46. We know from Figure 3a that the threshold between the strategies (1,1,1,1,1) and (3,2,0,0,0) is in this interval. In fact, there is only one threshold in our simple test network and five attacks. But when we investigate strategies that are not optimal we find same kind of phenomena as we discussed above. For example, when p_I is below 0.315 the second optimal strategy is (2,2,1,0,0) where a shift towards weaker links is occurs. In the simple test network the strategy (3,2,0,0,0) is already more optimal but in more complicated networks this kind of intermediate strategies may be optimal for some values of p_I.

To show that this can occur we look at Figure 5a where 13 attack strategies for the second test network of Figure 2a are shown for p_I between 0.48 and 0.64. Three thresholds are indicated with arrows for the optimal strategies (1,1,1,1,1,1,1,1,1,1,1,1,1), (2,1,0,1,0,2,0,1,1,1,2,1,1), (2,2,0,1,0,2,0,0,1,2,2,1,0) and (3,2,0,0,0,2,0,0,2,2,2,0,0) when p_I is in range (1, 0.622), (0.622, 0.588), (0.588, 0.510) and (0.510, 0.150) respectively. Between (0.050, 0.150) the strategy (3,2,0,0,0,3,0,0,2,2,1,0,0) is optimal and below 0.05 (3,3,0,0,0,3,0,0,1,3,0,0,0) is optimal. If we examine carefully low values of p_I we may find even more thresholds but they give same values for the metric in two or three decimals. In some applications, for example nuclear safety, small probabilities are important and exact results are needed for all values of p_I. Figure 5b shows optimal strategies for 26 attacks against the second test network. Targeted attacks shift from evenly distributed strategies towards weaker links, as before.

Finally, we note that there are many ways to extend the method of calculating the metrics where the link failure probabilities are unequal. Some examples are given below.

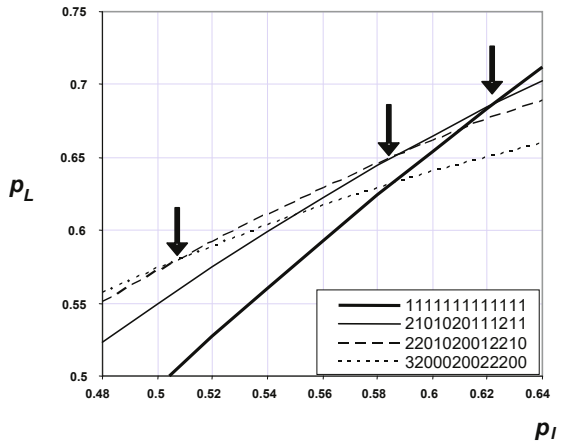

Figure 5a. The average metric of (5.4) for the second topology of Figure 3a is shown for the four strategies when p_I is between 0.48 and 0.64. All the strategies have 13 attacks against the network. The three changes of the optimal strategy for the attacker are indicated with arrows.

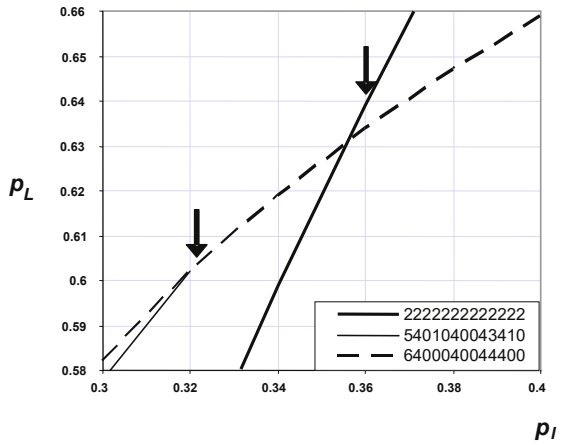

Figure 5b. The average metric of (5.4) for the second topology of Figure 3a is shown for the three strategies when p_I is between 0.3 and 0.4. All the strategies have 26 attacks against the network. The two changes of the optimal strategy for the attacker are indicated with arrows.

State 0 (line 1)	Node 4 or connection 1,4 failed, or both failed	Node 2 or connection 1,2 failed, or both failed
State 1 (line 2)	Node 4 and connection 1,4 operating	Node 2 or connection 1,2 failed, or both failed
State 2 (line 3)	Node 4 or connection 1,4 failed, or both failed	Node 2 and connection 1,2 operating
State 3 (line 4)	Node 4 and connection 1,4 operating	Node 2 and connection 1,2 operating

Table 1. States in matrix (6.2).

Attacks against the network may be conducted in wages and repairing of damaged links may occur between the wages. If damage against the network is observed before the next wave the network topology has changed—that is, some links are destroyed—and the new topology is input for the next calculation [4].

If hostile forces have more information about the systems in the network, for example, the probabilities for the nodes where command and control systems are, the two node metrics in the average metric of (5.4) should be weighted with these probabilities. In addition, more important and valuable nodes can be given more weight [8]. In the beginning of this paper we introduced the concepts of technical and functional nodes. Functional nodes have applications and, if we have this information only, we should include the functional nodes and the nodes where users connect the network in the metric. In this respect, a user is considered as a functional node.

If we study users in nodes i and applications in nodes j, (5.4) becomes:

$$P_L = \frac{1}{N_L(N_L-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{N_L} P_{F,j} W_{F,j} P_{L,ij} \tag{5.6}$$

where the probabilities for applications to be in node j is $P_{F,j}$ and the weighting factors for the value of the nodes are $W_{F,j}$.

6. NETWORKED COMMAND CENTRES

The last step is to combine the results of the previous sections. A networked system is composed of nodes and connections where failure probabilities for individual nodes and connections are denoted by p_N (notation p in Section 3) and $p_{L,ij}$ correspondingly. Consider a combination of one node and a connection between the node and a user in the network (second node is a user). The failure probability for the networked system is:

$$p_{N,ij} = p_{L,ij} + p_N - p_N p_{L,ij} \tag{6.1}$$

For the first example network of Figure 3a and the connection between the nodes one and four

$$p_{L,14} = 2p_l^2 + 2p_l^3 - 5p_l^4 + 2p_l^5$$

where p_l is the link failure probability. Usually, the failure probabilities are different for the networked system states, because of the network structure. The values of $p_{L,ij}$ are different on each line of the Markov matrix, for example, from (5.2):

$$p_{L,12} = p_l^2 + 2p_l^3 - 3p_l^4 + p_l^5$$

between nodes one and two and:

$$p_{L,23} = 4p_l^3 - 4p_l^4 + p_l^5$$

between nodes two and three.

As a simple example consider the Markov matrix where the time period is the approximate repair time as in Model 2 in Section 3:

$$\begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & p_1 & 1-p_1 \\ 0 & p_2 & 0 & 1-p_2 \\ p_1p_2 & (1-p_1)p_2 & p_1(1-p_2) & (1-p_1)(1-p_2) \end{vmatrix} \tag{6.2}$$

where $p_1 = p_{N=4,14}$ and $p_2 = p_{N=2,12}$ from (6.1). Node one is the user of the system. The model states are described in Table 1.

The stationary distribution of the example is:

$\pi_0 = p_1 p_2 \pi_3$,

$\pi_1 = p_2 \pi_3$,

$\pi_2 = p_1 \pi_3$ and

$\pi_3 = (1 + p_1 + p_2 + p_1 p_2)^{-1}$.

Generally, the metric for an entire network can be computed as:

$$M = \frac{1}{S} \sum_{j=1,\ldots,N} \sum_{s=1,\ldots,S} \pi_{js} = 1 - \frac{1}{S} \sum_{s=1,\ldots,S} \pi_{0s} \tag{6.3}$$

where S is the number of independent systems or applications in the network, N is the number of operational states in the system and state zero is the non operational state. The metric M takes values between zero and one where values near one indicate robustness of the networked system. The limiting probabilities $1-\pi_{0s}$, of Models 1 or 2, are used in (6.3). The failure probability p for a system of a node and a connection is given in (6.1). If users of the system connect to the network from different locations, the average over connection points should be included in (6.3).

Higher weights for critical applications may be used. Because weights for applications are difficult to evaluate, the metric M works better comparing alternate structures of a networked system. For one system s the value $M = 1-\pi_{0s}$ gives the metric which depends on the network topology of nodes and links.

If no repairing occurs, for example during targeted attacks, the metric can be computed for a specified time period with the methods of Section 5. From (5.4) the metric is $M = 1 - P_L$. Without repairing the stationary distribution gives the state where finally all the systems have failed and this is not useful information for comparing robustness of systems.

If the links are invulnerable the results of Section 3 are used and the network topology has no effect. In rare situations where the network is symmetric and symmetry is approximately maintained, $p_{N,ij} \approx p$ in (6.1) for all nodes and connections and the methods of Section 3 are suitable.

Model 1	Model 2B
$\pi = \dfrac{p_r + p_t}{\dfrac{p^{n-1}}{1-p^{n-1}} + (p_r + p_t)}$	$\pi = 1 - \dfrac{p^n}{(p_r + p)^n}$
$Q = \dfrac{\pi}{\pi_0} = \dfrac{(1-p)p_r + (1-p^{n-1})p_t}{p^n}$	$Q = \left(\dfrac{p_r}{p} + 1\right)^n - 1$
$n = \dfrac{\ln\left(\dfrac{p_r + p_t - pp_r}{pQ + p_t}\right)}{\ln p} + 1$	$n = \dfrac{\ln(Q+1)}{\ln(p_r + p) - \ln p}$

Table 2. Summary of Section 3.

The main results of Section 3 have been collected in Table 2. The first row is the probability for the system in operation, the second row gives the fraction Q of operation for n nodes and the third row gives the lower bound for the number of nodes if the capacity level Q has been given as a planning principle.

The steps of the method have been summarized here:

Step 1: Calculate or evaluate the failure probabilities for individual nodes and links. Firing theory or simulation can be used.

Step 2: (Section 3) Calculate the limiting distribution of failure probabilities for several replaceable nodes. Markov chains are used as a tool. (If more than one node is necessary for the system to operate then calculate the probability for the relevant combination of nodes by the same method.)

Step 3: (Sections 4 and 5) Calculate or evaluate the failure probabilities for connections of the system. One connection comprises of one or more links. (If more than one node is necessary for the system to operate then calculate the probability for the relevant combination of links by the same method.)

Step 4: (Section 6) Calculate the failure probabilities for pairs of a node and a link of the system. (If more than one node and one connection is necessary for the system to operate calculate the probability for the relevant combination of nodes and links by the same method.) In Step 4 each combination of a node and a link is considered as a 'combined node' and the new probability is inserted into the analysis of Step 2. Alternatively a specified time period can be investigated by the method of Section 5 if no repairing occurs or the assumption of stationary state distribution is not suitable. This may occur in targeted attacks of Section 5.

7. COMPARISON WITH RELATED WORK

Next we compare the methods and results of this work with the references given in Section 2. As a summary, in this paper we study random and targeted attacks against links of small networks where the measure of robustness is the average of operating connections over all the pairs of the network in (5.4). If several attacks are aimed at the same link the joint probability of the events is calculated from (5.5). A result of this paper is that the optimal attack is against all the network links when the probability of success is above a threshold. Below the threshold the attacks against links of nodes of lower degree are more optimal and usually this occurs via several intermediate thresholds for different configurations of links as explained in Section 5.

In [1,2] a symmetric network is suggested as a robust network against attacks. This is in good agreement with the results of this paper because the attacker is unable to take advantage of the degree distribution of the nodes what ever the probability of a successful attack is. A result of this paper is that if the number of attacks is high enough and the probability of success is above the threshold the network structure has no effect. If the number of attacks is less than the number links in the network the attacker should concentrate his limited attacks against the links of lower degree nodes.

In [4] different attack strategies have been examined and as a measure of robustness the size of the giant component or the average characteristic path length has been used because the ultimate aim is to disintegrate the network. The destruction probability is one for attacked nodes (edges) and when the fraction of nodes (edges) reaches the threshold value the network fails. The measure of robustness in this paper is the average connectivity of (5.4). Of course, if the information systems require connectivity to a large set of nodes the size of the giant component is a good measure. But this is a special case of (5.4) because of the comment given after (5.1) about three node systems (or more than three).

When using percolation theory it is assumed that the network is large and usually the measure of robustness is the size of the giant component. Despite the fact that percolation theory does not work for small networks the results in this paper are not in contradiction with the results of percolation theory. It is interesting to conduct a numerical exercise associated with the example of Figure 5 in [13] with the number of nodes N=100 and $\langle k \rangle$=2.1. The bimodal distribution given in Section 2 gives for r=0.1 the result rN=10 "hub" nodes of degree k_2=12 and for r=0.2 the result rN=20 "hub" nodes of average degree k_2=6.5. If we keep the sum constant p_t+p_r=0.11 we get p_r=0.1, p_t=0.01 and p_r=0.092, p_t=0.018 when r=0.1 and r=0.2 respectively. Thus if the fraction of targeted attacks against high degree nodes increases the number of the hubs increases. This makes the network more resilient against high degree node attacks because the network is more symmetric. In this paper the measure of robustness is different but we can make similar conclusions. For example in Figure 4a, if the link failure probability is below the threshold 0.45 it is optimal for an attacker to destroy the links of low degree nodes and as a counter measure the defender should plan his network more symmetric to increase the network robustness in (5.4). Above the threshold the attacker is superior and the network is vulnerable for all values of link failure probability. In this situation protection based on other means than the structure of the network is needed in order to lower the link failure probability. Optimal methods are selected by maximizing robustness and minimizing cost.

8. SUMMARY

The main results of this paper are the techniques of incorporating services, users, and the network topology in the same model and the calculation of targeted attacks against the network structure.

A networked system is studied in the network-centric warfare scenario. Failure probabilities for a system of several replaceable nodes have been calculated and a method for calculating failure probabilities for connections between nodes has been given.

A networked system is composed of nodes and connections where a connection has one or more links. The probabilities are for the worst case scenario, and as a baseline the networked system is assumed to be in operation, if only one node and one connection, or path, is in operation. The same model can be applied for systems of several nodes if the nodes are not replaceable. Markov chains and stationary system states have been used as a tool.

The theory presented in this paper gives the fraction of uptime $1-\pi_{0s}$ to downtime π_{0s} of the networked system *s*. In system design the lower bound for the number of replaceable nodes *n* can be computed when a specified ratio of uptime to downtime *Q* is required for the system. A new network metric can be computed as the normalized average of the values of $1-\pi_{0s}$ of Section 6. The metric can be computed for a system or for the entire network. In a network-centric warfare scenario, systems and networks can be compared using metric values.

Targeted attacks against nodes and links can be taken into account as described in Section 5. We show that for hostile forces to cause maximum damage, if locations of services in the network are unknown, is to attack all the network links for high values of link failure probability. For lower failure probabilities, below a threshold value, it is optimal for the attacker to shift the attacks against lower degree nodes. Usually the optimal set of links goes through several intermediate thresholds depending on the network structure. The threshold values of link failure probabilities for different combinations of links of the network are valuable information in planning and constructing robust network structures.

REFERENCES

[1] A.H. Dekker and B. Colbert, "Network Robustness and Graph Topology", *Proceedings of the 27th Australasian Computer Science Conference*, Dunedin, New Zealand, 26, 359–368, 2004. Available electronically at crpit.com/confpapers/CRPITV26Dekker.pdf.

[2] A.H. Dekker, "Network Topology and Military Performance", *MODSIM 2005 International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand, December 2005, pp. 2174–2180, 2005. http://www.mssanz.org.au/modsim05/papers/dekker.pdf.

[3] J. Galtier, A. Laugier, and P. Pons, "Algorithms to Evaluate the Reliability of a Network" *Proceedings 5th International Workshop on Design of Reliable Communication Networks*, 2005.

[4] P. Holme and B.J. Kim, "Attack Vulnerability of Complex Networks", *Physical Review*, E 65, 056106, 2002.

[5] S. Karlin and H. Taylor, *An Introduction to Stochastic Modelling*, Third Edition, Academic Press, 1998.

[6] M.L. Ling, "Proposed Network Centric Warfare Metrics: From Connectivity to the OODA Cycle", *MORS Journal*, Vol. 10 No. 1, April 2005.

[7] J. Moffat, *Command and Control in the Information Age Representing its Impact*, The Stationery Office, 2002.

[8] J. Moffat "Quantifying the Benefit of Collaboration Across an Information Network", *Journal of Defence Science*, Vol. 8, No. 3, pp. 123–129, September 2003.

[9] L.B. Page, "A Practical Implementation of the Factoring Theorem for Network Reliability", *IEEE Transactions on Reliability*, Vol. 37, No. 3, August 1988.

[10] L.B. Page and J.E. Perry, "Reliability Polynomials and Link Importance in Networks". *IEEE Transactions on Reliability*, Vol. 43, No. 1, March 1994.

[11] W.L. Perry and J. Moffat, *Information Sharing Among Military Headquarters—The Effects on Decision Making*, RAND Corporation, 2004

[12] P.S. Pulat, "Network Reliability with Arc Failures and Repairs" IEEE Transaction on Reliability, Vol. 37, No.3, 1988 August

[13] T. Tanizawa, G. Paul, R. Cohen, S. Havlin and H.E. Stanley, "Optimization of Network Robustness to Waves of Targeted and Random Attacks", *Physical Review*, E 71, 047101, 2005.

[14] J. Shaio, "A Family of Algorithms for Network Reliability Problems" *IEEE International Conference on Communications*, 2002.

[15] A.R. Washburn, *Notes on Firing Theory*, Naval Postgraduate School, 2002.

Vesa Kuikka works at the CIS Centre of the Finnish Defence Forces. He is a Licentiate of Philosophy (Physics) 1986 from the University of Helsinki. He also holds a Licentiate of Philosophy (Mathematics) degree from the Åbo Academy since 2004. Address Finnish Defence Forces, CIS Centre, Kutojankulma 2, PL 210, 02631 Espoo, Finland, e-mail vesa.kuikka@mil.fi.

VII

Optimal Sensor Placement in the Network Structure from a Defence Point of View

Vesa Kuikka and Juha-Pekka Nikkarila

Journal of Information Warfare, 19(2), 2020, pp. 46-61

Optimal Sensor Placement in the Network Structure from a Defence Point of View

V Kuikka, JP Nikkarila

Finnish Defence Research Agency
Riihimäki, Finland

E-mail: vesa.kuikka@mil.fi; juha-pekka.nikkarila@mil.fi

Abstract: *This article presents a method for optimising sensor placement in a general network topology. Sensor placement is calculated by balancing maximised closeness to the monitored nodes with sensor location distribution for protection. The model serves cyber defence planning and critical infrastructure defence. It uses node and link weights as parameters to describe information quantity, data importance to a defender, and criticality to the network operator. General network analysis, together with a penalty factor for describing distributed defence requirements, underpins the methodology. The Sprint operator network in the USA is used as an example to demonstrate the methodology.*

Keywords: *Network Resilience, Network Analysis, Cyber Defence, Optimal Sensor Placement, Network Sensors, Closed National Networks, National Segment of the Internet*

Introduction

This article presents a method for resolving the optimal placement of sensors in a general network topology. In this context, it defines a 'sensor' as a device or logic to monitor and control Internet traffic. The network model serves in planning cyber defence and defending critical information infrastructure. It includes node and link weights that can be applied to describe information quantity, data importance to a defender, or criticality to the network operator. Alternatively, the importance of data to an attacker may be considered. Most of the network models in the literature are 'local' and are therefore incapable of describing long-distance influence between nodes in a network. The model proposed in this article is 'global' (Kuikka 2018) considering all the paths in the network, making it possible to calculate the model's penalty components which represent the defensive capabilities of the model (Kuikka & Nikkarila 2019).

Network analysis, with an extra penalty parameter for describing distributed defence aspects, is used in the modelling. Penalty components are weighted with a tuneable parameter, and they are introduced to increase the defensive capability against targeted attacks. With low parameter values, sensors are placed in more central nodes close to each other. With high parameter values, sensors are placed on resilient locations, where sensors are more distributed in the network topology. The applicability of the method is illustrated with three real-world networks, and the results are both intuitive and practical. The same networks have been used in previous studies in analysing key features in the cyber capabilities of closed and open national networks. The Sprint operator network

in the United States is used as an example for comparing the authors' results with other research in the literature (Alenazi 2018). The placements of sensors in the Sprint operator network is demonstrated with both low- and high-parameter values.

The aspects of centrality and network resilience are considered equally. Adjusting both the link weights and the penalty parameter illustrates that the distribution of sensors can be either at the crucial crossroads in the network or widely separated from each other in the network structure. The parameter adjustment determines the level of protection. Resiliency is illustrated with a high parameter value when sensors are placed at more distributed locations. A well-placed arrangement of sensors is calculated by optimising centrality to monitored nodes and sensor protection capability. It is important to notice that, since the model proposed in this article is comprehensive as it takes into account all the possible variations, the computations may rapidly become heavy if the network grows or the number of desired sensors increases. However, it is then possible to utilise other computational methods for resolving exact or approximate solutions.

The cyber domain is a highly dynamic environment. For example, technical and procedural developments lead to new varieties of services. Consequently, systems are frequently reconfigured or replaced, and new software vulnerabilities are created and distributed continually. Defensive, offensive and cyber intelligence campaigns/operations are developed quickly. In dynamic environments such as this, good decisions require good situational awareness; see Endsley (1995a, 1995b) for example. As the cyber domain is dynamic, with frequently changing situations having cybersecurity implications, it is essential to develop fundamental defensive methodologies. It has been demonstrated that various defensive methodologies may have a role in the military context as well. One example of a military application is the closed national segment of the Internet established by Russia (Kukkola, Nikkarila & Ristolainen 2017b). This is presented in more detail in the background section.

Various types of sensors exist in modern communication network infrastructures and architectures. Sensors are utilised for both monitoring and controlling. Different technical solutions have been designed for cyber defence and traditional management functions. In cyber defence, sensor systems can be utilised to mitigate malicious actions and software or even hostile actors themselves. So-called 'fake news' represents a new tool that can be used for influencing public opinion or for engaging in hybrid warfare. Structural development of cyberspace and national segments of the Internet change a nation's capabilities of situational awareness.

In a concurrent study (Nikkarila & Kuikka 2019), the authors demonstrated how complex network analysis can be utilised for evaluating the effects of asymmetry established by national networks. They recognise that there are several possibilities to improve the cyber defence of open national networks. For example, it is possible to utilise probabilistic models of war, proposed in the literature (Cioffi-Revilla 1989; Cioffi-Revilla & Dacey 1989; Nikkarila *et al.* 2018). The current research presents one possibility to enhance the defensive capability of an open national network.

This article presents a method for determining optimal sensor placement in a communication network at both the physical and logical levels. Physical topology is a map of network structure, and logical topology illustrates how services are used within the network. The method is useful in planning cyber defence and defending information systems. The model has node and link weights

as parameters to describe the value of data and services in the nodes and information transfer via the links. Sensors are devices or software logic to monitor and control communication traffic. Network analysis with a phenomenological penalty parameter to describe the distributed defence requirements is used to calculate the optimal placement of sensors in the network structure. The penalty parameter is a free parameter and its numerical value depends on how much protection is needed. The value depends on link and node weights, although the dependency is not very strong.

The article further demonstrates the usability of the method using illustrations from three real-world networks. One of the networks has been used in a previous article by the authors in analysing the differences in operational cyber capabilities between closed and open national networks; this article further develops the understanding of these networks.

The Sprint operator network in the United States has been used for examples in several previous studies. When compared with these studies, the authors' results are closest to the *k*-median results (Alenazi 2018). This example entails placing sensors in the Sprint operator network with low and high values in the penalty parameter. With a low parameter value, sensors are placed in central nodes as close as possible to each other. Additionally, this presents a more resilient arrangement with high parameter values, where sensors are placed in more distributed locations. Both aspects of centrality and network resilience are covered in a well-balanced manner.

Background: Changes in Cyberspace and the Closing of National Infrastructure Networks

As cyberspace is a synthetic system-of-systems concept, it is worthwhile to consider that these systems can be further modified by organisations and nation-states. Some of the modifications are understandable and acceptable while others have elements of an arms race. One potential (and non-desirable) outcome is a fragmented global network.

In the summer of 2016, when NATO recognised cyberspace as a military domain, Russia nearly concurrently declared that RuNet—the Russian segment of the Internet—would have a capability to be disconnected from the global Internet by 2020 (Ristolainen 2017). This move aimed to shape cyberspace to gain a military advantage in a potential conflict. Russia's potential objectives at the tactical level and the effects of the process were studied by Nikkarila and Ristolainen (2017).

Subsequently, Kukkola, Nikkarila, and Ristolainen (2017a, 2017b) demonstrated that these kinds of structural changes in cyberspace would have substantial and widespread implications and could generate inequality between two competitors. In the predictive analysis, this process was assessed to lead towards the formation of national segments of cyberspace. This would create new types of cyber threats against the remaining 'open network society'. The current understanding is that Russia aims to establish digital sovereignty by 2024 (Kukkola, Nikkarila & Ristolainen 2019). The situational awareness of a decision-maker in a closed national segment of the Internet (Kukkola 2018b) may be superior to that of its counterpart which uses a national network and relies on open networks (Kukkola, Nikkarila & Ristolainen 2019).

Since the 2016 declaration, Russia has continued the network 'closing' process and other nations (for example, China, North Korea, and Iran) may be developing their versions of closed

national segments of the Internet. Even more nations may follow their examples. The scale of the closing process is important since it may cause substantial effects at the strategic level as well (Kukkola, Nikkarila & Ristolainen 2017a, 2017b). Consequently, it is important to monitor potential followers and their alliances. Nevertheless, it is important to take into account various potential developments to be prepared for different future scenarios. Also, technical and procedural solutions of the closing process have been studied in Kukkola (2018b, 2018c). The role of the national segment of the Internet as a part of the system-of-systems of cyber defence has been examined in Kukkola (2018a).

Various mathematical methods have been used to demonstrate the effects of a closing process on the cyber capabilities of open and closed parties (Nikkarila *et al.* 2018; Nikkarila & Kuikka 2019). Additionally, wargaming has been proposed as a tool to illustrate the effects of the closing process at the strategic level (Lantto, Åkesson *et al.* 2018). Effects on nations that implement closed national segments can be predicted. Resilience and confrontation scenarios can be predicted in cases where some nations have implemented a closed national segment of the Internet whereas others have not (Lantto, Huopio *et al*. 2018). A simple, two-party wargaming scenario is shown in **Figure 1**, below, to illustrate the analytical process. Siukonen *et al.* (2019) present a methodology to determine the status of a national network.

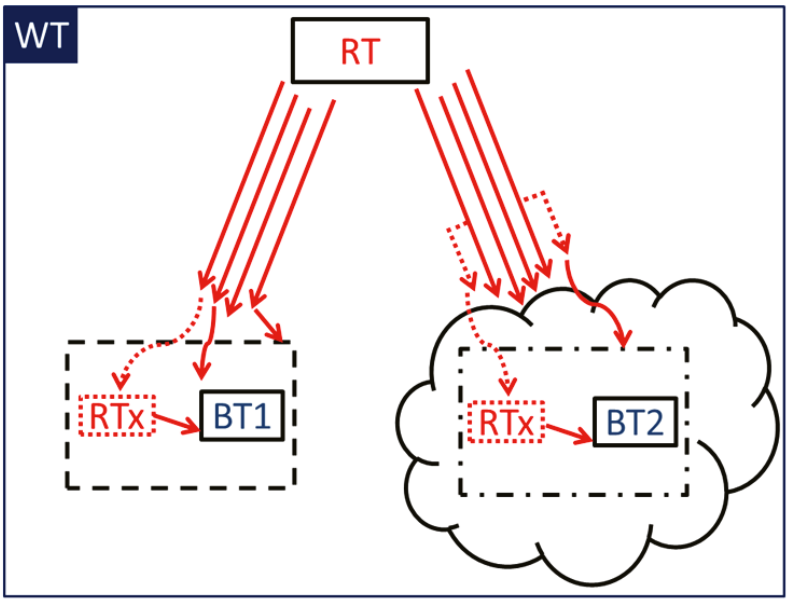

Figure 1: Schematic outline of the two-sided cyber wargame

In the illustration in **Figure 1**, the open network is presented on the left. Blue Team 1 (BT1) relies on the open network infrastructure (dashed line box). The closed national network is presented on the right. Blue Team 2 (BT2) relies on technologically different network infrastructure (dashed-dotted line box). The Red Team (RT) is actively attacking both blue teams and possibly operating inside both networks (RTx). The arrows represent attack methods and their variations. Attacks can be targeted both against the infrastructure or the specific blue team systems. The control team (White Team [WT]) possesses all the knowledge in the game and initiates all the events. Therefore, the WT is visualised as a solid line encompassing everything in the game (Lantto, Huopio *et al*. 2018).

Methods

Methods of complex networks analysis described by Cherifi *et al*. (2018) and network propagation modelling described by Kuikka (2018) is used to study communication networks. The first sample network is a simple version of an existing real-world infrastructure network. Because one goal of this study is to introduce modelling methods, a small network is appropriate for illustrating the methodology.

To model the process of accessing networked services, this illustration uses the information-spreading model proposed in Kuikka (2018) and Kuikka and Nikkarila (2019). This model is suitable for describing communication networks because connecting network services is typically a step-by-step process in a network structure from a source node to a target node. The model has no explicit relaying functionality. However, network traffic produced by sensors' activities can be modelled with bi-directional node and link weights. This closely resembles the classical connectivity theory dealing with probabilities of connection between the nodes in a network (Colbourn 1987; Kuikka 2019; Kuikka & Syrjänen 2019).

In the methodology of complex networks and social network analysis, the method for optimal sensor placement can be based on two alternative measures: closeness centrality or betweenness centrality. Closeness centrality indicates how close a node is to all other nodes in the network. Nodes with high betweenness have considerable influence within a network by their control over information passing between others. The two measures for optimal sensor placements are the corresponding sums of the influence measures of nodes with an extra parameter where the influence between the sensors is eliminated. The extra component is multiplied by a tunable parameter that acts as a penalty element in the model. The penalty parameter is incorporated in the model to increase the capability to defend against targeted attacks. The network operator can strengthen the distributed defence by increasing the parameter value. It is crucial that the model is 'global' and takes into account all the paths in the network to make it possible to calculate the penalty component.

The proposed network analysis method considers all the possible self-avoiding paths in a network. A self-avoiding path is a sequence of steps on a path that do not visit the same node more than once. The inclusion of only self-avoiding paths describes information transmission between a source and a target node. The model also considers information flow direction and both link and node weights. The weights can be applied to model information quantity and data importance to a defender or criticality to the network operator.

Weights describe the resilience of nodes and links in the network structure from the viewpoint of the process or phenomenon under interest. Two different aspects are considered—the monitored or controlled process and the resilience of the network. Moreover, there are numerous possible metrics for measuring the resilience of networks (Lü *et al*. 2016). Consequently, it is left up to user discretion to plan and select weights. The authors provide several examples that help in selecting weight values based on these principles. Fortunately, in many cases, a wide range of weights yields similar sensor placement in the network topology. In other words, the placement of sensors is not highly sensitive to link or node weights.

The mathematical methodology is briefly presented here to highlight the main features of the model. More detailed calculations and a pseudo algorithm were presented in Kuikka (2018). The mathematical model considers all paths between nodes in the network. The model considers the common links before the branching of the paths. Paths from a source node to a target node are combined iteratively in the descending order of common path lengths (number of links) at the beginning of their paths. Practical examples have been provided in previous research (Kuikka 2018). Node and link weighting factors are implemented in the model by including all the corresponding weighting values along a path. These are denoted by W_L in the equation below. The equation shows how two paths with path lengths L_1 and L_2 and a common path length of L_3 are combined.

$$P_{i,\max(L_1,L_2)}(T) = W_{L_1} D_{L_1}(T) + P_{i-1,L_2}(T) - \frac{W_{L_1} D_{L_1}(T) P_{i-1,L_2}(T)}{W_{L_2} D_{L_2}},$$

$$\text{where } i = 1, \dots, N_L - 1 \ \ (L_1, L_2 \leq L_{max}),$$

$$P_{0,L_2} = W_{L_2} D_{L_2}(T).$$

In this equation, $P_{i-1,L_2}(T)$ is the intermediate result at step i during the iterations, and N_L is the number of different paths between the source node s and the target node t. The temporal distribution as a function of time T of the spreading process (Kuikka 2018) is denoted by $D_L(T)$. For larger networks, an upper limiting value for L_{max} is necessary to limit computing time. In a later step, $P_{i-1,L_2}(T)$ is used as an input on the right side of the equation when shorter common path lengths, in turn, are combined during the computation. Finally, all N_L paths go into the result $P_{N_L-1}(T)$ for the probability of propagation from the source node to the target node via all possibble paths shorter or equal to L_{max}. The limiting value $\lim_{T\to\infty} P_{N_L-1}(T)$ is particularly interesting, as the time scales of information transmission are short when compared with the time horizon of the analysis. The limiting value of the temporal distribution is $\lim_{T\to\infty} D_L(T) = 1$ in the above equations. The procedure provides the probability $C_{s,t}$ for spreading from node s to node t. The authors' approach uses the following expression for the measure of sensor placement:

$$M(\pi) = \sum_{\substack{s\notin S \text{ and } t\in S \text{ or} \\ s\in S \text{ and } t\notin S}} C_{s,t} - \pi \sum_{s\in S \text{ and } t\in S} C_{s,t}.$$

The first summation (component) is taken over sensor/non-sensor pairs of indexes and the second summation over sensor pairs of indexes. The set of nodes equipped with a sensor are denoted by S. In the first summation, index pairs between non-sensor nodes and between sensor nodes are not included. The second part is the penalty component multiplied by the penalty parameter π. The measure M(π) is maximised in the procedure for determining the optimal locations of the desired number of sensors. Later in this article, all the numerical results are calculated by M(π).

The first component in $M(\pi)$ is based on the closeness centrality measure defined in Equation (11) in Kuikka (2018). As mentioned earlier, an alternative measure M_b is based on the betweenness measure of Equation (14) in Kuikka (2018):

$$M_b(\pi_b) = \sum_{n \in S} b_n - \pi_b \sum_{s \in S \text{ and } t \in S} C_{s,t},$$

$$\text{where } b_n = \frac{C - B_n}{C} \text{ and } C = \sum_{s,t \in \text{All nodes}} C_{s,t}.$$

The quantity b_n is the betweenness centrality value of node n. The numerical value of B_n is equivalent to the closeness centrality measure when node n is removed from the network. It is to be expected that in most cases the optimal values of measures $M(\pi)$ and $M_b(\pi_b)$ are close to each other because normally central nodes have also a high betweenness value. In other words, these quantities are highly correlated. For atypical network topologies or system configurations, the two objective functions can provide different results. A semantic difference between the quantities is that $M(\pi)$ measures access to information, and $M_b(\pi_b)$ emphasises information flow between the nodes in the network. These points of view may be different in planning backbone infrastructure networks, local networks, or embedded system structures.

In the literature, most of the network models include only local interactions; for this reason, they are incapable of describing long-distance influence between nodes. The network model in this article includes all interactions between the nodes of a network, that is short- and long-distance connections, making the calculation of the penalty component possible in $M(\pi)$ and $M_b(\pi_b)$. Penalty components are weighted with tuneable penalty parameters, like π and π_b. These parameters are introduced to increase the defensive capability against technical network failures or targeted attacks.

Example Networks

Three different real-world network topologies are used for demonstrating the model. As an introduction, a model with a simple physical network structure is presented. The physical network layer is only one part of analysing the 'big picture'. Additionally, the use of service layers on top of the physical network with the simple network (**Figure 2**, below) is illustrated. To prove the usability of the methodology in a more realistic situation, where the number of sensors is higher, calculations are performed with a well-known network. Therefore, as the third example, the Sprint operator network in the United States is analysed.

The same examples are applied in earlier studies or the literature. The first network was applied in Nikkarila & Kuikka (2019) where the authors analysed the differences in operational cyber capabilities between closed and open national networks. The second network describes the same network as the first network, but the complex structure is shown in more detail. The Sprint operator network has been studied in Alenazi (2018), where the sensor placements have been determined with four different models. The results from those models can be compared with the model in this article.

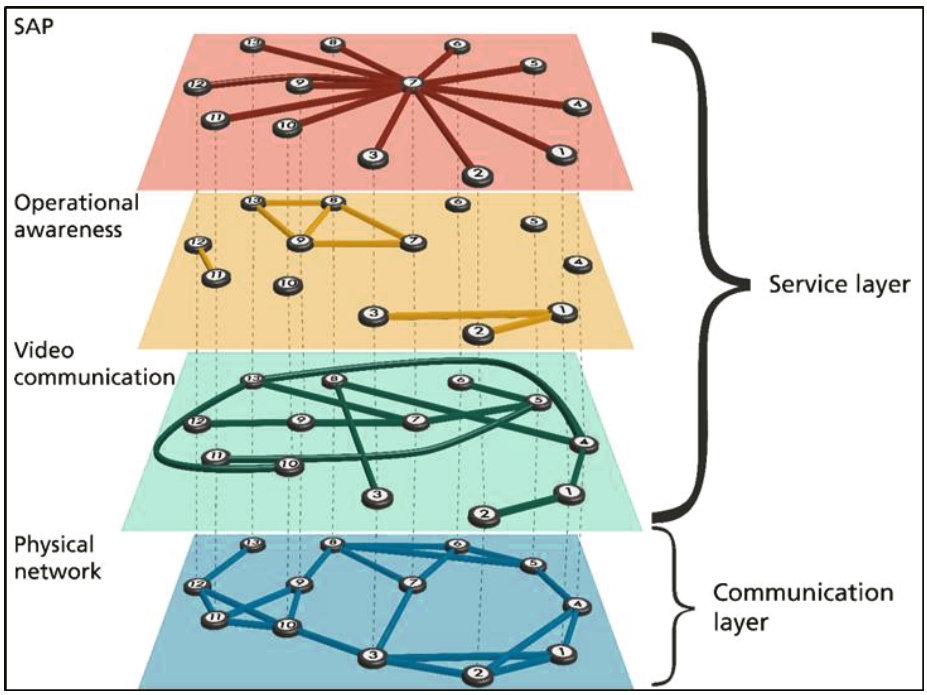

Figure 2: Networked service layers on the physical network layer. The physical topology is a map of network structure, and the logical topology illustrates how services are used within the network (Kuikka & Syrjänen 2019).

Services are essential from a user's point of view. Consequently, the sensor placement should consider not just the physical network structure, but also the networked services. Services can be categorised and rated based on the value of cyber and operational functionalities. In a military context, capability is based on accessibility and usability of these functionalities. In real-world applications, the value of services depends on the task and mission of the scenario. Therefore, the values are also time and user dependent (Kuikka & Syrjänen 2019).

Results

The first example network is a simplified version of a real-world infrastructure network. This network is selected for illustrating the method of selecting optimal sensor placements. Three parameter values are to be specified in the modelling: the penalty parameter, link weights, and the number of sensors to be deployed. In the spreading model, a limiting value of time approaching infinity is describing the equilibrium state. Because self-avoiding paths are assumed to describe information transmission in communication networks, the number of visits per node in a path is limited to one.

Table 1, below, shows an analysis of the network topology of **Figure 3**, also below, which comprises the 'communication layer' in **Figure 2**, above. Some negative values of placing sensors are indicated in the table. These cases are regarded as statistical outliers. The previous configurations 1, 6, 11, 13 and 8, 12, 13 in **Table 1** may be outliers. Higher link weights

with a high number of sensors are extreme cases. The essential question is which model parameter values should be used in the calculations. The number of sensors is determined by the managers or operators of the network. The decision depends on the cost of the technology and the number of available resources for maintaining the sensor system.

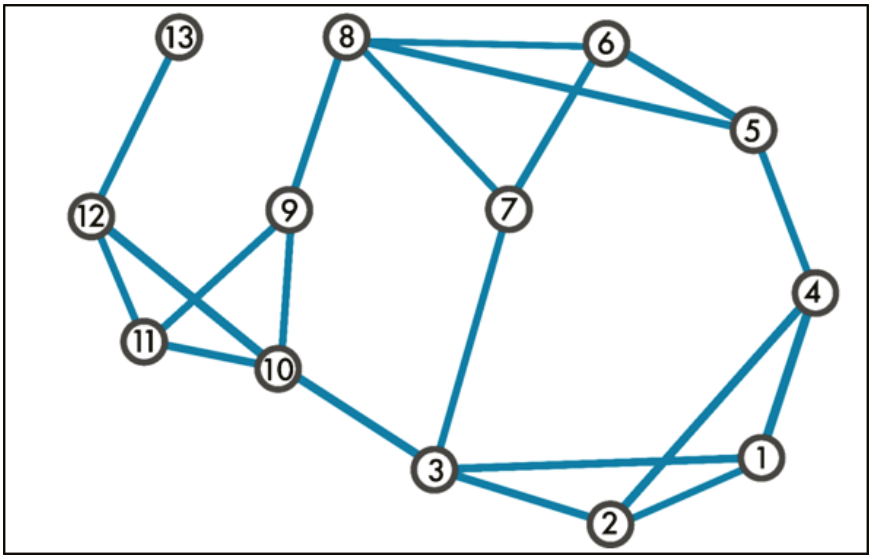

Figure 3: Example network of 13 nodes and 20 bidirectional links (40 links). Communicating devices are modelled as nodes, and connections between the devices are modelled as links between the nodes.

The optimal placement of sensors has a clear pattern with a wide range of model parameters. This is a favourable feature of the model because results are not highly sensitive to the model parameters. For example, it may be difficult to estimate the best numerical value for the link weights. Very coarse estimations and experimenting with a few link weight values are sufficient.

The results in **Table 1**, below, show that, when only one sensor is placed in the network, node 3 is optimal. Further, nodes 3 and 8 are optimal for two sensors. Different configurations for three sensors are suggested by the analysis: nodes 1, 8, and 10 are optimal for low activity events with low requirements for protection and 1, 8, and 12 for high protection. The corresponding configurations are 3, 8, and 10 (1, 6, and 13) for moderate activity events with low (high) protection. The last columns show the results for a very wide range of managed and monitored events in the network traffic. The final selection can be made among the suggested nodes, calculated with different model parameters, and judging with possible additional budgetary and system management requirements.

The second example is a real-world communication network represented in **Figure 4**, below. The network has some characteristic structures of larger communication network topologies. Typically, networks have few nodes with many links or many nodes with few links. This has implications in optimal sensor placement because the degree that a node affects the possibilities for sensors to detect network traffic and events in data transmission.

W_l	π	#	Sensors	W_l	π	#	Sensors	W_l	π	#	Sensor
0.1	1	1	3	0.5	1	1	3	0.9	1	1	3
		2	3,8			2	3,8			2	3,8
		3	1,8,10			3	3,8,10			3	3,7,8
		4	3,4,8,11			4	1,6,8,10			4	3,6,7,8
		5	3,4,6,9,12			5	1,3,6,8,11			5	3,6,7,8,10
	5	1	3		5	1	3		5	1	3
		2	3,8			2	4,11			2	3,8
		3	1,8,12			3	1,6,13			3	8,12,13
		4	3,4,8,12			4	1,6,11,13			4	negative

Table 1: Optimal sensor placements computed for the network topology of **Figure 3**. Links weights, penalty parameter values, the number of sensors, and optimal nodes are denoted by W_l,π,#, and sensors correspondingly

The larger network topology of **Figure 4**, below, requires more sensors than the network in **Figure 3**. Additionally, the penalty parameter describing distributed protection of sensors should have a higher numerical value. For the analysis, it is useful to calculate sensor placement with different model parameter values.

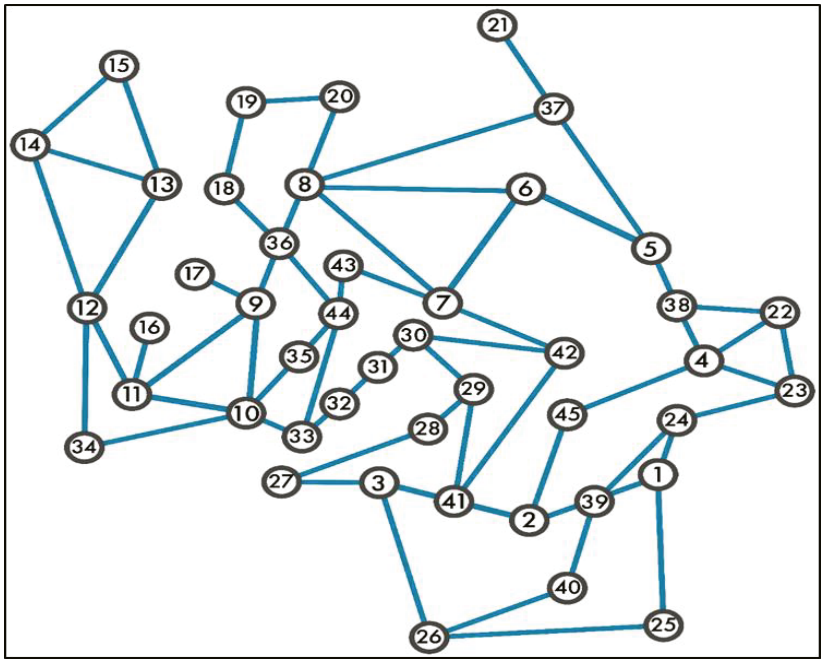

Figure 4: Example network of 45 nodes and 64 bidirectional links (128 links)

Table 2, below, shows the analysis with different model parameter values. The link weight values describe monitoring low activity events and describe more probable activities of the network. The penalty parameter values describe moderate distributed protection and high protection requirements of sensors. The choice can be made according to these planning guidelines. In practice, other factors are taken into account which makes the decision easier. Financial resources, physical environments, and other resources should be optimised. Consequently, the analysis can be used to develop a possible set of sensor locations.

W_l	π	#	Sensors		W_l	π	#	Sensors
0.1	5	1	8		0.5	5	1	8
		2	8,10				2	8,10
		3	8,10,41				3	8,10,39
		4	4,8,10,41				4	8,10,22,41
		5	4,8,10,39,41				5	3,4,8,10,13
		6	4,8,10,14,39,41				6	1,8,10,13,22,29
		7	3,4,8,10,14,30,39				7	1, 8, 10,15,22,27,31
		8	3,4,8,10,14,30,39,44				8	1,10,15,19,21,22,27,31
	10	1	8			10	1	8
		2	8,10				2	10,39
		3	8,10,39				3	8,14,39
		4	4,8,10,41				4	8,13,24,28
		5	4,8,10,29,39				5	1,15,21,28,44
		6	4,8,14,17,29,39				6	1,14,17,19,21,28
		7	3,4,8,10,14,30,39				7	14,16,19,21,22,25,28
		8	3,4,8,10,14,18,30,39				8	negative

Table 2: Placement of sensors in the network topology of **Figure 4** with different model parameters

The final design of the configuration is determined by considering other aspects of planning and real-world constraints.

Table 3, below, shows an example of sensor placement in the network topology of **Figure 3,** above, where the service layers of **Figure 2**, above, are also considered. The results in **Table 1** (four columns), above, are different because services carry a high weight in the calculations in **Table 3**. If the physical network layer or the availability of services was more important, there would be more results similar to those in **Table 3**. For example, **Figure 1**, above, shows that SAP services from node 7 to all users in the network are important (Kuikka & Syrjänen 2019).

W_l	π	#	Sensors		W_l	π	#	Sensors
0.1	1	1	7		0.1	5	1	7
		2	7,11				2	1,9
		3	1,9,13				3	1,9,13
		4	1,5,9,13				4	1,5,9,13
		5	1,5,8,9,13					

Table 3: Placement of sensors in the network topology of **Figure 3** and service layers of **Figure 2** (Kuikka & Syrjänen 2019)

Figure 5, below, shows the placement of five sensors in the Sprint operator network with differing link weight and penalty parameter values (in the figure: 0.1 the uppermost row, 0.5 the middle row, and 0.9 the lowest row) and the penalty parameter value of 1 (left column) and 100 (right column). Optimal sensor placements are calculated based on closeness centrality, self-avoiding paths, and maximum path lengths of 75. This study indicates that increasing the link weight centralises the optimal sensor locations. On the other hand, increasing the penalty parameter value yields more protection through optimised sensor placement. Therefore, a higher penalty parameter value leads to a more distributed configuration of sensors. Interestingly, with the highest weight value (lowest row), the results are reasonable only at a narrow region of the

penalty parameter values and this phenomenon is not clearly seen with the highest link weight. Consequently, in those cases, the sensors are placed in central locations and may then be close to each other. In some environments, a centralised sensor placement is acceptable, yet it may increase vulnerability to spontaneous network failures or cyberattacks.

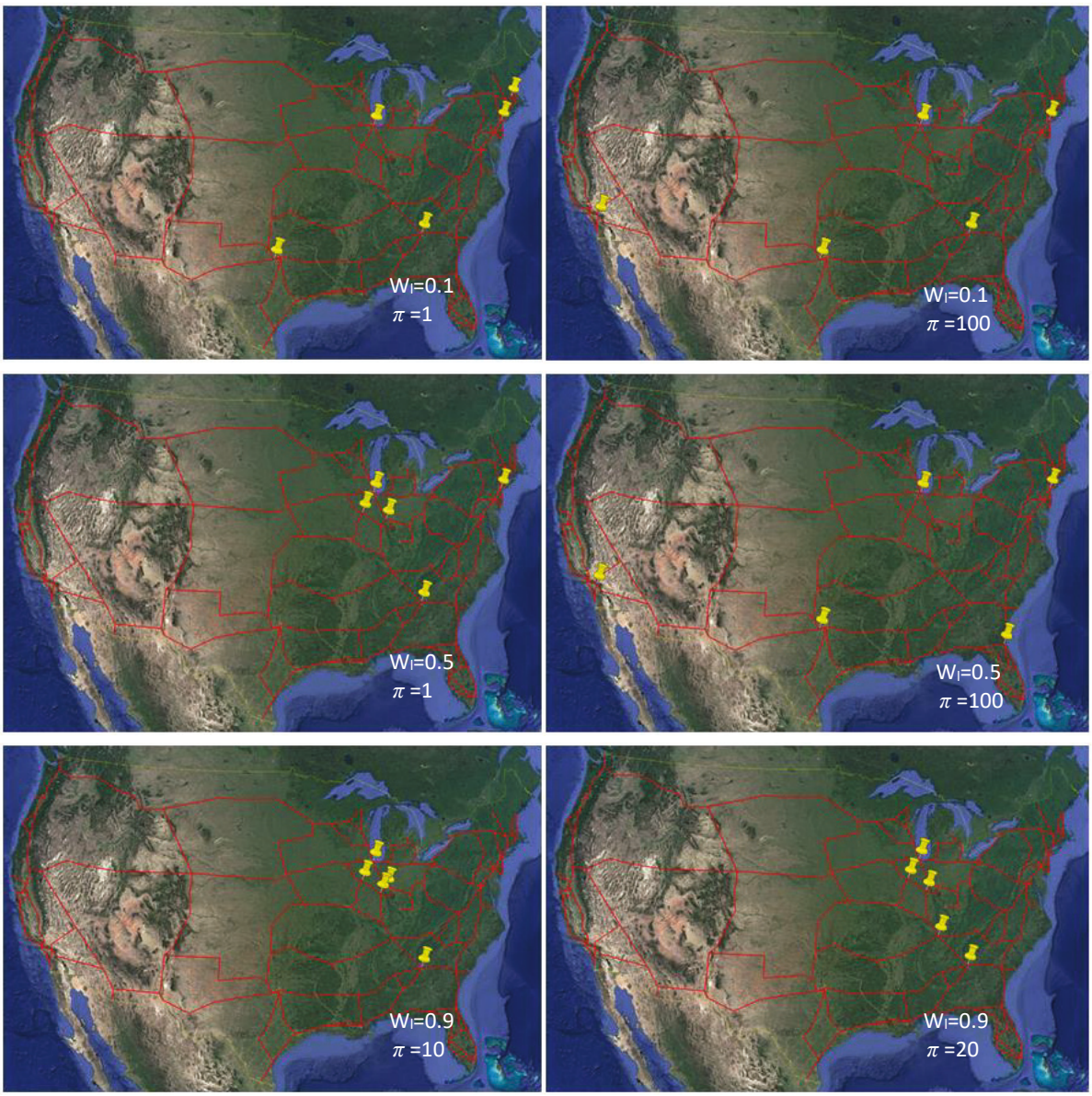

Figure 5: Optimal locations of five sensors in the Sprint network (Google Earth) with different penalty parameter values (π) and link weight values (W_l). Protection capability for a specified W_l is achieved by increasing the numerical value of π.

It is also noteworthy that results with parameter values W_l = 0.5, π = 10 correspond very accurately with *k*-median results, according to Alenazi (2018). According to the study, sensor placement with the *k*-median principle minimises the overall latency in the network. The authors' model does not take into account the actual propagation time of the Internet traffic but instead considers the topology of the network. Even so, the model may lead to comparable sensor placement. To conclude, the authors have proposed a model that can be used for a distributed sensor placement in an optimised manner. Both aspects, closeness centrality and network resilience, are taken into account in a balanced way. It could be worthwhile to study how resilient a network with a sensor placement calculated by the model is when compared to optimised networks, for example, AC=Algebraic Connectivity, NC=Network Criticality, or *k*-median (Alenazi 2018).

By selecting the number of sensors as in Alenazi (2018), it is possible to directly compare the results. Algebraic connectivity and network criticality have been used as robustness functions in Alenazi (2018). The sensor placements described in this article are different from these two methods, as can be seen in Alenazi (2018). However, the sensor placements for the *k*-median results are almost similar. Comparing different methods would require selecting which processes are monitored and/ or controlled (both can be done at the same time) and selecting criteria for evaluating the resilience of networks (Lü *et al*. 2016).

Conclusions

This article introduces a sensor placement method that considers monitoring and controlling processes, resiliency of the network, networked services, and structure of the entire network. This article proposes a method for optimising sensor placement in a complex network topology. Optimal placement of sensors is important when a limited number of sensors are positioned in the communication network infrastructure. The optimisation calculation of sensor placement accounts for all paths between nodes in the network. The novelty of the methodology is that network structures, the value of information services, the importance of network connections, and protection aspects through distributed sensor placements are considered consistently in the same model. An additional value of the proposed model is that it can be utilised when improving defensive cyber capabilities in a nationwide context.

The article presents a model incorporating real-world communication networks and demonstrates that the results are intuitive and practical. It compares the results with earlier studies conducted with the Sprint network in the United States. When compared with earlier studies, the authors' results with the *k*-median (Alenazi 2018) seem to be the closest approximation.

Requirements for protection are accomplished by implementing a penalty parameter to distribute sensors farther away from each other in the network structure. Adjusting the parameter value determines the level of protection. A balanced configuration of sensor placements is computed by optimising closeness centrality to monitored nodes and optimising the protection capability of the sensors. The proposed model can assist when constructing a large-scale, real-time cyber situation awareness system.

This article has presented a method for calculating optimal sensor placement in a network topology. Sensors are placed on optimal and resilient locations to monitor network traffic or control activities between network segments, network devices, information services, or systems' users.

In the basic model, the network structure is described with only two types of elements: nodes and links. In more detailed models, different kinds of objects or services with their interrelationships can be investigated. Additionally, sensors with different functionalities could be modelled as a system of systems. Complex network analysis and systems engineering methods provide powerful tools for future research in this area.

References

Alenazi, MJF 2018, 'On SDN controller placement to achieve robustness against targeted attacks', *Proceedings of the 6th International Conference on Complex Networks & Their Applications, Complex Networks 2017, Studies in Computational Intelligence,* vol. 689, eds. C. Cherifi, H, Cherifi, M Karsai & M Musolesi, Springer, Cham, CH.

Cherifi, C, Cherifi, H, Karsai, M & Musolesi, M (eds.) 2018, *Proceedings of the 6th International Conference on Complex Networks & Their Applications, Complex Networks 2017, Studies in Computational Intelligence,* vol. 689, Springer, Cham, CH.

Cioffi-Revilla, C 1989, 'Mathematical contributions to the scientific understanding of war', *Mathematical and Computer Modelling,* vol. 12, nos. 4-5, pp. 561-75.

——and Dacey, R 1989, 'The probability of war in then-crises problem: Modeling new alternatives to Wright's solution', *Synthese,* vol. 76, no. 2, pp. 285-305.

Colbourn, CJ 1987, *The combinatorics of network reliability*, Oxford University Press, Oxford, UK.

Endsley, MR 1995a, 'Toward a theory of situation awareness in dynamic systems', *The Journal of the Human Factors and Ergonomics Society (HUM FACTORS)*, vol. 37, pp. 32-64.

——1995b, 'Measurement of situation awareness in dynamic systems', *The Journal of the Human Factors and Ergonomics Society (HUM FACTORS),* vol. 37, pp. 65-84.

Kuikka, V 2018, 'Influence spreading model used to analyse social networks and detect sub-communities, *Computational Social Networks*, vol. 5, no. 12, <https://doi.org/10.1186/ s40649-018-0060-z>.

——2019, 'Modeling network resilience and utility of services', *IEEE Systems Conference*, IEEE, Orlando, FL, USA, <https://doi.org/10.1109/SYSCON.2019.8836780>.

——& Nikkarila, JP 2019, 'Optimal sensor placement in network topology from the defence point of view', *Proceedings of the 18th European Conference on Cyber Warfare and Security*, ACPI, eds. T Cruz & P Simoes, University of Coimbra, PT, 4-5 July 2019, pp. 248-58.

——& Syrjänen, M 2019, 'Modelling utility of services in military networked environments', *Proceedings of the 2019 International Conference on Military Communications and Information Systems (ICMCIS)*, Budva, ME, 2019, pp 1-6.

Kukkola, J 2018a, 'The Russian segment of Internet as a resilient battlefield', *ISMS Annual Conference 2018, Military Sciences and Future Security Challenges*, Warsaw, PL.

——2018b, 'Civilian and military information infrastructure and the control of the Russian segment of Internet', *International Conference on Cyber Warfare and Security (ICCWS) 2018,* Wash-ington, DC, US, 8-9 March 2018.

——2018c, 'Russian cyber power and structural asymmetry', *Proceedings of the 13th Internation-al Conference on Cyber Warfare and Security (ICCWS),* National Defence University, Washington DC, US, 8-9 March, pp. 362-8.

——, Nikkarila, JP & Ristolainen, M 2017a, 'Shaping cyberspace – A predictive analysis of adversarial cyber capabilities', *IST-145 specialists' Meeting predictive analytics and analysis in the cyber domain,* Sibiu, RO.

——2017b, *Game changer: Structural transformation of cyberspace*, Riihimäki: Finnish Defence Research Agency, viewed 14 May 2018, <http://puolustusvoimat.fi/web/tutkimus/tutkimuslaitok-sen-julkaisut>.

——2019, *Game player: Facing the structural transformation of cyberspace*, Riihimäki: Finnish Defence Research Agency, viewed 13 September 2019, <http://puolustusvoimat.fi/web/tutkimus/tutkimuslaitoksen-julkaisut>.

Lantto, H, Åkesson, B, Kukkola, J, Nikkarila, JP & Ristolainen, M 2018, 'Wargaming a closed na-tional network: What are you willing to sacrifice?', *Proceedings of MILCOM 2018,* Los Angeles, CA, US, pp. 29-31.

Lantto, H, Huopio, S, Åkesson, B, Nikkarila, JP, Suojanen, M, Ristolainen, M & Tuukkanen, T 2018, 'Wargaming the cyber resilience of structurally and technologically different networks', *ISMS Annual Conference 2018: Military Sciences and Future Security Challenges,* Warsaw, PO.

Lü, L, Chen, D, Ren, XL, Zhang, QM, Zhang, YC & Zhou, T, 2016, 'Vital nodes identification in complex networks', *Physics Reports*, vol. 650, pp. 1–63.

Nikkarila, JP & Kuikka, V 2019, 'Complex network analysis for evaluating effects of asymmetry established by closed national networks', *International Conference on Cyber Warfare and Security (ICCWS) 2020,* Norfolk, VA, US, 12-13 March 2020.

Nikkarila, JP & Ristolainen, M 2017, '"RuNet 2020"- Deploying traditional elements of combat power in cyberspace?', *Proceedings of 2017 International Conference on Military Communications and Information Systems (ICMCIS),* 15-16 May 2017, pp. 1-8.

Nikkarila, JP, Åkesson, B, Kuikka, V & Hämäläinen, J 2018, 'Modelling closed national networks – Effects in cyber operation capabilities', *Proceedings of the 17th European Conference on Cyber Warfare and Security (ECCWS),* Oslo, NO, pp. 248-56.

Siukonen, V, Ristolainen, M, Nikkarila, JP & Kukkola, J, 2019, 'Are we there yet?, Monitoring the technical deployment of the national segments of the Internet', forthcoming, *Proceedings of MILCOM 2019,* November 2019, Norfolk, VA, US.

National Defence University
P.O. Box 7, 00861 HELSINKI

Tel. +358 299 800

www.mpkk.fi

ISBN 978-951-25-3194-3 (Pbk.)
ISBN 978-951-25-3195-0 (Pdf)
ISSN 2342-9992 (print)
ISSN 2343-0001 (web)

Puolustusvoimat
The Finnish Defence Forces