The Case for Campaign Analysis: A Method for Studying Military Operations

Rachel Tecott   Andrew Halterman

19 October 2020

Abstract

Military operations, from armored warfare to nuclear counterforce attacks to humanitarian intervention, are fundamental to international relations theory and practice. Although scholars have studied military operations for decades, there is little methodological guidance to direct their analysis or to help them address broader theoretical questions. Campaign analysis is a method that involves the use of a model and techniques for managing uncertainty to answer questions about military operations. In this article, we define campaign analysis, standardize the method, provide methodological guidance, and illustrate the promise of the method for academic theory. We identify six core steps of the method: 1) question selection, 2) scenario development, 3) model construction, 4) value assignment, 5) sensitivity analysis, and 6) interpretation and presentation of results. Additionally, we recommend that scholars elevate the models they build in their analyses as a central contribution of their work, and we recommend a new technique for propagating uncertainty in inputs through to a model's output. We then conduct replications and extensions of two existing campaign analyses in order to illustrate the six steps of campaign analysis, the value of the two recommendations, and the promise of campaign analysis as a technique for improved measurement of variables central to international relations theories.

Rachel Tecott is a PhD Candidate in Political Science at MIT and a member of MIT's Security Studies Program. For the 2020-2021 academic year, she is a pre-doctoral fellow at the Institute for Security and Conflict Studies (ISCS) at the Elliott School of International Affairs at George Washington University. Andrew Halterman is a PhD Candidate in Political Science at MIT and a member of MIT's Security Studies Program and Political Methodology Lab.

The authors would like to thank Eric Heginbotham, Tim McDonnell, Aidan Milliff, Asfandyar Mir, Reid Pauly, Barry Posen, Brad Potter, Daryl Press, seminar participants at MIT, panelists at APSA, and three anonymous reviewers for helpful comments.

Could NATO thwart a Warsaw Pact armored invasion?[1] Could the United States destroy Russia's nuclear forces in a first strike?[2] Could Iran disrupt the flow of oil from Saudi Arabia or close the Strait of Hormuz?[3] Why did the Coalition achieve such staggering success in the Gulf War?[4] Would China be able to distinguish between conventional escalation and strategic counterforce in a conflict with the United States over Taiwan?[5] These are questions of tremendous policy importance. Their answers shape doctrine, guide procurement, inform force posture, and influence decisions to go to war. Military operations are also central to scholarship in international relations, where the likely outcomes of hypothetical wars are central variables in core theoretical debates. These questions are difficult to tackle due to the complexity and uncertainty of combat. Most efforts to answer them occur within governments, think tanks, and government-funded research centers, but scholars have an important role to play.

Campaign analysis is a method involving the use of a model and techniques for managing uncertainty to answer questions about military operations. The method involves six steps: formulating a question, specifying a political scenario, constructing a model that represents the military operation, setting values for those variables using qualitative research and technical military information, running the model with sensitivity analysis, and interpreting the output of the model and presenting the conclusions of the analysis.[6]

Security studies is in the midst of a period of methodological innovation. A variety of research techniques used by security studies scholars have recently been formalized as methods with guidance on when and how to employ them. Archival research, long a cornerstone of international security research, has recently received careful attention as a method of inference.[7] Wargaming has been reexamined by researchers as a method for data generation and theory testing.[8] Researchers have also developed sophisticated methods for measuring core security studies concepts, such as territorial control and perception of threats and

[1] John J Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe," *International Security*, Vol. 7, No. 1 (1982), pp. 3–39, doi:10.2307/2538686; Joshua M Epstein, *Measuring Military Power: The Soviet Air Threat to Europe*, vol. 199 (Princeton University Press, 1984); Barry R Posen, "Measuring the European Conventional Balance: Coping with Complexity in Threat Assessment," *International Security*, Vol. 9, No. 3 (1984), pp. 47–88, doi:10.2307/2538587; Joshua M Epstein, "Dynamic Analysis and the Conventional Balance in Europe," *International Security*, Vol. 12, No. 4 (1988), pp. 154–165, doi:10.2307/2538999; Barry R Posen, "Is NATO Decisively Outnumbered?" *International Security*, Vol. 12, No. 4 (1988), pp. 186–202, doi:10.2307/2539002.

[2] Keir A Lieber and Daryl G Press, "The End of MAD? The Nuclear Dimension of US Primacy," *International Security*, Vol. 30, No. 4 (2006), pp. 7–44, doi:10.1162/isec.2006.30.4.7.

[3] Joshua R Itzkowitz Shifrinson and Miranda Priebe, "A Crude Threat: The Limits of an Iranian Missile Campaign Against Saudi Arabian Oil," *International Security*, Vol. 36, No. 1 (2011), pp. 167–201, doi:10.1162/ISEC_a_00048; Caitlin Talmadge, "Closing Time: Assessing the Iranian Threat to the Strait of Hormuz," *International Security*, Vol. 33, No. 1 (2008), pp. 82–117, doi:10.1162/isec.2008.33.1.82; Eugene Gholz and Daryl G Press, "Protecting 'the Prize': Oil and the US National Interest," *Security Studies*, Vol. 19, No. 3 (2010), pp. 453–485, doi:10.1080/09636412.2010.505865.

[4] Stephen Biddle, "Victory Misunderstood: What the Gulf War Tells Us About the Future of Conflict," *International Security*, Vol. 21, No. 2 (1996), pp. 139–179, doi:10.2307/2539073.

[5] Caitlin Talmadge, "Would China Go Nuclear? Assessing the Risk of Chinese Nuclear Escalation in a Conventional War with the United States," *International Security*, Vol. 41, No. 4 (2017), pp. 50–92, doi:10.1162/ISEC_a_00274.

[6] We use the term "campaign analysis" because it is the term most scholars currently use to describe the method we formalize here.

[7] Christopher Darnton, "Archives and Inference: Documentary Evidence in Case Study Research and the Debate over US Entry into World War II," *International Security*, Vol. 42, No. 3 (2018), pp. 84–126, doi:10.1162/ISEC_a_00306.

[8] Erik Lin-Greenberg, "Game of Drones: The Effect of Remote Warfighting Technology on Conflict Escalation (Evidence from Wargames)," *Available at SSRN 3288988*; Reid BC Pauly, "Would US Leaders Push the Button?

signals.[9]

Campaign analysis has not yet received such treatment. Although scholars have used the method for decades, campaign analysis remains underspecified and techniques for conducting it are an oral tradition among a small number of scholars.[10] This paper defines, standardizes, and provides guidance on how to employ the method of campaign analysis.

Scholars can use campaign analysis both to inform policy and to advance academic debate. Scholarly campaign analysis provides an important independent counterweight to government views on issues of national and international consequence. If the marketplace of ideas is to function properly, then multiple, rigorous analyses must be presented.[11] Academics can employ campaign analysis to offer independent assessments of the sufficiency of a force posture, how a possible attack could unfold, or which factors are most likely to affect the costs of conflict. Academic researchers can serve the public interest, as they do in other areas of policy, by bringing their substantive knowledge, research design skills, and independence to bear to inform public discourse around military operations.

Campaign analysis can also serve as a powerful tool of measurement for theories of international relations that have often relied on imperfect proxy variables. Theories of deterrence, coercion, and credibility often require estimating the ability of states to use their militaries to get what they want, while many theories of misperception, bureaucratic politics, and domestic politics are concerned with establishing gaps between actual and perceived military power. The theory of the nuclear revolution,[12] for example, hinges in large part on the ability of states to threaten adversaries' retaliatory capabilities, as do contemporary theories of nuclear coercion. Offense-defense theory depends on a measurement of ease of conquest, but scholars have relied on a variety of problematic proxies ranging from the technology across a dyad to "bean counts" of military forces. These measures fail to capture the concept at the core of the matter: the ease with which one side would be able to accomplish a military objective vis-a-vis the other side.

Campaign analysis offers an alternative approach to measuring these variables that takes into account not only a state's resources but how it would employ them in a specific sce-

---

Wargames and the Sources of Nuclear Restraint," *International Security*, Vol. 43, No. 2 (2018), pp. 151–192, doi:10.1162/isec_a_00333; Elizabeth M Bartels, "Building Better Games for National Security Policy Analysis" (PhD thesis, Pardee RAND Graduate School, 2020).

[9]Therese Anders et al., "Measuring Territorial Control in Civil Wars Using Hidden Markov Models: A Data Informatics-Based Approach," *arXiv Preprint arXiv:1711.06786*; Marika Landau-Wells, "Dealing with Danger: Threat Perception and Policy Preferences" (PhD thesis, Massachusetts Institute of Technology, 2018); Azusa Katagiri and Eric Min, "The Credibility of Public and Private Signals: A Document-Based Approach," *American Political Science Review*, Vol. 113, No. 1 (2019), pp. 156–172, doi:10.1017/S0003055418000643.

[10]Researchers engaged in campaign analysis have included some methodological guidance in their articles: Epstein discusses the use of a model and management of uncertainty to answer questions about hypothetical military operations, Mearsheimer gives high-level guidance on conducting a "theater balance", and Kupchan argues for careful approaches to quantification, sensitivity analysis, and thinking about clear "yardsticks of sufficiency" in analyzing military operations. No previous work, however, has defined the method of campaign analysis, standardized it, or provided comprehensive guidance for how to conduct and evaluate it. Epstein, *Measuring Military Power*. John J Mearsheimer, "Numbers, Strategy, and the European Balance," *International Security*, Vol. 12, No. 4 (1988), pp. 174–185, doi:10.2307/2539001. Charles A Kupchan, "Setting Conventional Force Requirements: Roughly Right or Precisely Wrong?" *World Politics*, Vol. 41, No. 4 (1989), pp. 536–578, doi:10.2307/2010529. Alain C Enthoven and K Wayne Smith, *How Much Is Enough?: Shaping the Defense Program, 1961-1969* (Santa Monica: RAND Corporation, 1971).

[11]We thank an anonymous reviewer for their framing of this point.

[12]Brodie; Robert Jervis, *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon* (Cornell University Press, 1989).

nario for a specific objective against an opponent with its own capabilities and objectives. While the number of US, Russian, and Chinese nuclear warheads is alone too crude a proxy for nuclear survivability, campaign analysis can account for improved US intelligence and warhead accuracy and Russian decisions about how to deploy mobile missiles and submarines to reveal shifts in the United States' ability to threaten Russia's retaliatory capability. Campaign analysis equips the researcher with the tools to estimate the factors that offense-defense theory suggests drives or discourages aggression, such as the ratio of attacker casualties to defender casualties (the loss-exchange ratio), attacker casualties per square kilometer of territory conquered, or the probability of attacker victory, all in a particular, plausible operation. Campaign analysis thus offers enormous promise as the most precise available measurement tool for variables at the center of international relations theory and practice.

Critics have raised several objections to campaign analysis. Scholars have argued that campaign analysis oversimplifies, leaving out political-economic-social variables and focusing only on one or a handful out of the many possible scenarios that could shape conflict.[13] This critique misses the purpose of campaign analysis. A study that sheds light on the predicted outcome of one carefully specified military operation can be important in its own right, regardless of whether the same study addresses the wider range of political pathways that could lead conflict to unfold in different directions. Moreover, as Posen observes, specialized attention to individual scenarios is a necessary step towards larger, cumulative questions about the overall balance of military power between states: "analysis is about dividing problems into their component parts to permit focused, specialized attention to the parts."[14]

Campaign analysis has also been criticized on the grounds that academic security studies researchers do not have the resources or (classified) information to conduct technical analysis of military operations.[15] More broadly, war could simply be too complex and uncertain to model.[16] These are important critiques. Military operations are difficult to model well, and many critical variables cannot be estimated with precision. Academic researchers, however, are not only familiar with military operations and the political conditions that shape them, but they are also equipped with the principles of good research design necessary for the rigorous management of complexity and uncertainty. The method of campaign analysis that we formalize in this paper is explicitly designed to facilitate valid inference in the face of uncertainty. In particular, campaign analysis manages uncertainty through the development and use of a model–a transparent description of relationships between inputs and output–and careful parameter value assignment. While campaign analysis does not enable researchers to perfectly explain or predict every aspect of military operations, if well done, the method does equip researchers to answer specific questions of great value to policy and academic research.

The purpose of this article is to encourage an increase in the quantity and quality of aca-

[13]Eliot A Cohen, "Toward Better Net Assessment: Rethinking the European Conventional Balance," *International Security*, Vol. 13, No. 1 (1988), pp. 50–89, doi:10.2307/2538896.

[14]Posen in John J Mearsheimer, Barry R Posen, and Eliot A Cohen, "Correspondence: Reassessing Net Assessment," *International Security*, Vol. 13, No. 4 (1989), pp. 128–179, doi:10.2307/2538782, p. 146

[15]Cohen, "Toward Better Net Assessment," especially pp. 59-60, 85.

[16]Carl von Clausewitz, *On War*, Translated and Edited by Michael Howard and Peter Paret (Princeton University Press, 1976); Barry D Watts, "Ignoring Reality: Problems of Theory and Evidence in Security Studies," *Security Studies*, Vol. 7, No. 2 (1997), pp. 115–171, doi:10.1080/09636419708429344.

demic campaign analysis. In service of this objective, this article provides a methodological toolkit for researchers and illustrates the value of the method for policy and theory. The rest of this article is divided into five sections. First, we distinguish campaign analysis from related methods of military science and establish its boundaries and scope. Second, we elaborate on the definition of campaign analysis, standardize the six core steps of the method, and propose methodological guidance for valid inference at every step. Third, we propose two recommendations for improving the method: clear presentation of the core model used in a campaign analysis to improve the transparency and usability of campaign analyses and the use of an "input distribution" approach and Monte Carlo techniques to better handle uncertainty. Fourth, we replicate and extend two published campaign analyses–Wu's analysis of Chinese nuclear survivability and Posen's analysis of NATO's prospects against the Warsaw Pact. In these replications, we illustrate the six steps of campaign analysis, the benefits of our two recommendations, and the value of the method for international relations theory. We conclude with summary and discussion of future campaign analysis research.

Campaign Analysis - A Distinct Method

The method of campaign analysis took root in academic security studies in the 1980s in the context of efforts to assess the outcome of hypothetical armored conflict in Western Europe.[17] Since then, researchers have applied the technique of campaign analysis to a wide range of engagements including nuclear operations,[18] air strikes or missile attacks,[19] irregular war and humanitarian operations,[20] operations at either the theater or tactical level,[21] and analysis of historical rather than hypothetical campaigns.[22]. Some types of questions,

[17] Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe."; Epstein, *Measuring Military Power*; Posen, "Measuring the European Conventional Balance."; Epstein, "Dynamic Analysis and the Conventional Balance in Europe."; Barry R Posen, *Inadvertent Escalation: Conventional War and Nuclear Risks* (Cornell University Press, 1991). Researchers in academic security studies had been using similar techniques since the 1970s to study nuclear conflict, though: Lynn Etheridge Davis and Warner R Schilling, "All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask," *Journal of Conflict Resolution*, Vol. 17, No. 2 (1973), pp. 207–242, doi:10.1177/002200277301700203; John D Steinbruner and Thomas M Garwin, "Strategic Vulnerability: The Balance Between Prudence and Paranoia," *International Security*, pp. 138–181, doi:10.2307/2538581.

[18] Lieber and Press, "The End of MAD?"

[19] Whitney Raas and Austin Long, "Osirak Redux? Assessing Israeli Capabilities to Destroy Iranian Nuclear Facilities," *International Security*, Vol. 31, No. 4 (2007), pp. 7–33, doi:10.1162/isec.2007.31.4.7; Shifrinson and Priebe, "A Crude Threat."

[20] Kelly M Greenhill, "Mission Impossible? Preventing Deadly Conflict in the African Great Lakes Region," *Security Studies*, Vol. 11, No. 1 (2001), pp. 77–124, doi:10.1080/714005314; Alan J Kuperman, *The Limits of Humanitarian Intervention: Genocide in Rwanda* (Brookings Institution Press, 2004); Bjoern H Seibert, "African Adventure?: Assessing the European Union's Military Intervention in Chad and the Central African Republic" (MIT Security Studies Program, 2007)

[21] Although we use the term "campaign analysis," the method does not exclude tactical or operational questions. See "DOD Dictionary of Military and Associated Terms," https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf, p. 161. for a discussion of different levels of warfare and Wayne P Hughes, "Overview," in Wayne P Hughes, ed., *Military Modeling for Decision Making* (Military Operations Research, 1989) for an alternative taxonomy.

[22] Brian McCue, *U-Boats in the Bay of Biscay: An Essay in Operations Analysis* (National Defense University Press, 1990); Michael J Armstrong and Michael B Powell, "A Stochastic Salvo Model Analysis of the Battle of the Coral Sea," *Military Operations Research*, pp. 27–37; Michael J Armstrong and Steven E Sodergren, "Refighting Pickett's Charge: Mathematical Modeling of the Civil War Battlefield," *Social Science Quarterly*, Vol. 96, No. 4 (2015), pp. 1153–1168, doi:10.1111/ssqu.12178; Niall MacKay, Christopher Price, and A Jamie Wood, "Weighing the Fog of War: Illustrating the Power of Bayesian Methods for Historical Analysis Through the Battle of the Dogger Bank," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, Vol. 49, No. 2 (2016), pp. 80–91,

such as those relating to counterinsurgency, may involve more uncertainty and difficulty in modeling than a single-move missile strike campaign, but nothing in the definition of campaign analysis excludes campaigns with greater uncertainty. Likewise, historical operations have much less uncertainty than hypothetical operations.The availability of historical data to inform their study of the engagement means that researchers do not need to rely on a model and can use familiar techniques such as process tracing.[23] In historical cases, though, the use of a model can help answer counterfactual questions about the operation.

Campaign analysis is one of many methods that academic, government, and military researchers have developed to understand conflict.[24] Some techniques, including wargames, tabletop exercises, and field exercises stretch back thousands of years.[25] Others took root during WWII and over the course of the Cold War. These methods include the (sometimes overlapping) fields of operations research, systems analysis, modeling and simulations, and net assessment.

During World War II, the Allied governments began conducting operations research, drawing on applied mathematics to improve tactical and operational force employment decisions. In particular, operations research informed Allied antisubmarine warfare, bombing techniques, and submarine tactics.[26] Scholars conducting operations research today often focus on abstract approaches to solving general classes of problems, rather than examining specific military scenarios with real-world data. Military operations research tends to address optimization problems such as how to optimally allocate search efforts when looking for a target,[27] or how to optimally allocate warheads to targets,[28] while the broader operations research literature (in computer science and business schools) has studied a wide range of constrained optimization problems.[29]

Closely related to operations research, "systems analysis" and "modeling and simulations" methods develop models to assist military commanders with a wide range of applied tasks, including in planning and predicting the outcomes of specific operations. Many of these models combine components from operations research with data from field testing of equip-

doi:10.1080/01615440.2015.1072071; Brennen Fagan et al., "Bootstrapping the Battle of Britain," *Journal of Military History*, Vol. 84, No. 1 (2020); Ryan T Baker, "Logistics and Military Power: Tooth, Tail, and Territory in Conventional Military Conflict" (PhD thesis, George Washington University, 2020)

[23]Stephen Biddle, "Speed Kills? Reassessing the Role of Speed, Precision, and Situation Awareness in the Fall of Saddam," *Journal of Strategic Studies*, Vol. 30, No. 1 (2007), pp. 3–46, doi:10.1080/01402390701210749, p. 8

[24]The term "method" does not have a strict definition in social science research and is often used interchangeably with the terms "research design," "technique," or "tool." See, for example, Darnton, "Archives and Inference.".

[25]Paul K Davis, "Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation," *Proceedings of the IEEE*, Vol. 83, No. 8 (1995), pp. 1138–1155, doi:10.1109/5.400454; Roger D Smith, "Essential Techniques for Military Modeling and Simulation," in, *1998 Winter Simulation Conference. Proceedings (Cat. No. 98CH36274)*, vol. 1 (IEEE, 1998), pp. 805–812, doi:10.1109/WSC.1998.745067.

[26]Philip McCord Morse and George E Kimball, *Methods of Operations Research*, 1951, ed. (Courier Corporation, 2003).

[27]Henry R Richardson and Lawrence D Stone, "Operations Analysis During the Underwater Search for Scorpion," *Naval Research Logistics Quarterly*, Vol. 18, No. 2 (1971), pp. 141–157, doi:10.1002/nav.3800180202; Lawrence D Stone, *Theory of Optimal Search* (Elsevier, 1976); Lawrence D Stone et al., *Optimal Search for Moving Targets* (Springer, 2016).

[28]Ravindra K Ahuja et al., "Exact and Heuristic Algorithms for the Weapon-Target Assignment Problem," *Operations Research*, Vol. 55, No. 6 (2007), pp. 1136–1146, doi:10.1287/opre.1070.0440.

[29]For more recent general military operations research techniques, see Mike Cornforth and Wayne P. Jr. Hughes, *Military Modeling for Decision Making*, 3d. ed (Military Operations Research Society, 1999); Alan R Washburn and Moshe Kress, *Combat Modeling*, vol. 139 (Springer, 2009). For applied civilian operations research, see the efficiency of the single checkout line at Trader Joe's.

ment, expert opinion, and historical experience, and often integrate many sub-models into larger, multi-resolution models. The most sophisticated models, such as the STORM or RSAS models used in the US Defense Department, can help commanders plan for everything from the tactical outcomes of armored battles in a specific piece of real world terrain to the projected use of spare parts during air operations.[30]

Campaign analysis evolved from and remains closely related to operations research, systems analysis, and modeling and simulations methods. Most campaign analyses begin with a specific, substantive question about a particular military operation with strategic implications. This distinguishes campaign analysis from ideal-type operations research, which tends to prize solutions to general classes of abstract problems, and from military models and simulations, which are often designed so that many users with very different questions can use the same tool. For instance, a Defense Department model or simulation could produce answers both for division commanders who are interested in losses and kills in an engagement, and also for maintenance personnel who need a better grasp of spare parts requirements.[31] The complexity of some modeling and sims provide great benefit in studying the detailed components of military operations, but their complexity should not dissuade scholars from constructing simple models, which have their own advantages. Because researchers employing campaign analysis tend to ask only one or several specific questions under a limited number of scenarios, they can use simpler models than the models required to answer many potential questions at different layers of resolution.

Campaign analysis is also related to yet distinct from game theory and formal modeling. Game theory, the study of rational decision-making in strategic interactions, was also used during World War II[32] and fundamentally shaped defense thinking and scholarship during the Cold War.[33] Though campaign analysis and formal modeling share an effort to make causal assumptions explicit, transparent, and often formal, formal modeling in social science tends to model human decisions in abstracted, logically derived games, whereas campaign analysis tends to model the outcome of military operations while explicitly controlling for specified human (political) decisions.

Finally, campaign analysis should not be confused with the much larger, holistic project of net assessment."Net assessment" refers to the collection of concepts and techniques originally pioneered by Andrew Marshall to help the US government plan for long-term competition with Soviet Union. Institutionalized in the Department of Defense's Office of Net Assessment, the approach is characterized by an emphasis on long-term trendlines, analysis of the adversary, and attention to the myriad ways that political, economic, and social factors shape military competition and hypothetical military engagements across a wide range of future scenarios.[34] Net assessment is thus an expansive research agenda.

In contrast, researchers use campaign analysis to tackle much more discrete projects. Rather

---

[30] For a history of Defense Department modeling and simulation efforts see Davis, "Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation."

[31] E.g. TACWAR. Robert J Atwell and D Graham McBryde, "Theater-Level Ground Combat Analyses and the Tacwar Submodels" (Institute for Defense Analyses, Alexandria, VA, 1991).

[32] Morse and Kimball, *Methods of Operations Research*; Oskar Morgenstern and John Von Neumann, *Theory of Games and Economic Behavior* (Princeton university press, 1953).

[33] See, perhaps most prominently, research on nuclear deterrence including the seminal Thomas C Schelling, *Arms and Influence* (Yale University Press, 1966).

[34] Thomas G. Mahnken (ed), *Net Assessment and Military Strategy: Retrospective and Prospective Essays* (Cambria, 2020).

than assess long-term trends, analyze a wide variety of scenarios, and examine how fluctuations in all political, economic, and social variables could shape the future of military competition and conflict, campaign analyses focus on answering a single question in the context of a carefully specified scenario. Confusion between the two approaches, exacerbated by the absence of any shared standards for campaign analysis as a distinct methodology, gave rise to methodological debate in the 1980s in *International Security*. Substantive debate between NATO "pessimists" and NATO "optimists" evolved into discussion of the feasibility and value of academic campaign analysis. Cohen, with a background in net assessment, criticized Posen and Mearsheimer for their narrow focus on specific scenarios and for omitting political variables from their analysis. Posen and Mearsheimer countered that they had not intended to embark on the more expansive project, and had deliberately focused their analysis around narrower questions about specific military operations under specific conditions. Although Posen and Mearsheimer used the term "net assessment" to describe their work, they were in fact articulating a fundamental distinction between net assessment and what we call campaign analysis, and defending the feasibility and utility of the method.[35]

Six Steps of Campaign Analysis

Campaign analysis is a method that involves the use of a model and techniques for managing uncertainty to answer questions about military operations. In this section, we standardize the six core steps of campaign analysis and provide guidance for how to conduct them for valid inference. Researchers conducting campaign analysis 1) formulate a question with a precise outcome, 2) define a political scenario, 3) construct and defend a model that represents the military operation within that scenario and outputs the outcome that answers the question, 4) sets values for model parameters, 5) runs the model with sensitivity analysis, and 6) interprets and presents the results of the analysis. Although we call them steps, the process of conducting campaign analysis is often an iterative process. We provide specific guidance on how to conduct each step.

Step 1: Question and Outcome

Campaign analysis, like all methods in social science, is intended to answer motivating research questions. Campaign analysis thus requires the familiar social science process of transforming broad, motivating questions into narrower, more concrete, well-specified questions with precise and measurable outcomes.

Most published academic campaign analyses are predictive exercises, answering questions about the likely outcome of consequential military operations. For instance, could Iran de-

[35]Other researchers have also used the term "net assessment" to refer to methods that fit our definition of campaign analysis. For instance, Glaser and Kaufmann's definition of net assessment is much closer to what we call campaign analysis: "analyses of the ability of a country's forces to perform military missions against the forces of an opponent." Charles L Glaser and Chairn Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?" *International Security*, Vol. 22, No. 4 (1998), pp. 44–82, doi:10.1162/isec.22.4.44, p. 74. See also Keir A Lieber, "Mission Impossible: Measuring the Offense-Defense Balance with Military Net Assessment," *Security Studies*, Vol. 20, No. 3 (2011), pp. 451–459, doi:10.1080/09636412.2011.599193; Noel Anderson, "Peacekeepers Fighting a Counterinsurgency Campaign: A Net Assessment of the African Union Mission in Somalia," *Studies in Conflict & Terrorism*, Vol. 37, No. 11 (2014), pp. 936–958, doi:10.1080/1057610X.2014.952260.

stroy Saudi oil refineries or close the Strait of Hormuz?[36] Could NATO conventional forces defend Western Europe against a Soviet attack?[37] These analyses shed light on inherently important military campaigns. These questions are often framed as "sufficiency" questions: is a campaign achievable, impossible, or is the outcome indeterminate? Glaser and Kauffman call this approach the "adequacy of particular force postures." In other words, "Are our existing forces sufficient to defeat this contingency? If not, would this alternative force be sufficient?"[38] Researchers can also use campaign analysis to examine the effects of a specific variable on the outcome of hypothetical military operation. For instance, Wu asks how the level of dispersion of Chinese nuclear forces changes their survivability.[39]

How should researchers select a question? As in the rest of social science, there are no methodological rules to guide question selection.[40] Broadly speaking, researchers tend to choose questions based on interest and to reduce uncertainty enough to answer their question. In the past, most researchers seem to have chosen questions for their policy importance. Fewer academics have used campaign analysis to inform broader theoretical academic debates.[41] Other questions might be amenable to the method of campaign analysis, but may not be interesting to social science researchers. For instance, highly technical questions such as the best equipment for soldiers to carry in their packs on counterinsurgency patrols are unlikely to interest social science researchers. Other questions are uninteresting because they are so implausible that the answer does not teach us something interesting (e.g. how a Costa Rican nuclear capability would change Central American politics).

Previous researchers have carefully chosen their questions to limit uncertainty in the model and parameters. Researchers may select questions for which only a few variables really matter, so they can answer their question with a simple model and a manageable amount of research. If researchers discover in the research process that the question would require a model with more variables than they can manage, they may end up abandoning the questions. Researchers may also choose questions in which something important can be learned in an initial interaction. Sometimes the first move of a campaign is decisive for how the campaign turns out: nuclear first strikes and territorial *faits accompli* are examples. Questions that require analysis of many interactions are more difficult to study than short or single move operations, as each decision point creates a fork in the path to an outcome. The more interactions an operation has, the more difficult it will be for the researcher to answer the question with enough confidence to make the effort worthwhile.

A different source of uncertainty concerns the value of variables within the model and affects the questions that campaign analysis researchers pursue. For instance, we might be interested in how the B-2 stealth bomber affected the US ability to deliver nuclear weapons, but if we lack (classified) information on the B-2's radar detectability, we cannot answer a question about an operation driven by that variable. Although knowledge of parameter values is rarely perfect, researchers may choose to avoid questions where there is close to no publicly available information to inform value assignment for critical variables.

---

[36] Shifrinson and Priebe, "A Crude Threat"; Talmadge, "Closing Time."

[37] Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe"; Posen, "Is NATO Decisively Outnumbered?"

[38] Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?", p. 75.

[39] Riqiang Wu, "Living with Uncertainty: Modeling China's Nuclear Survivability," *International Security*, Vol. 44, No. 4 (2020), pp. 84–118, doi:10.1162/isec_a_00376.

[40] Stephen Van Evera, *Guide to Methods for Students of Political Science* (Cornell University Press, 1997).

[41] See, as one exception, Wu, "Living with Uncertainty."

From Question to Outcome

Researchers conducting campaign analysis then move from motivating questions to specific outcomes of interest. A general question, such as whether the United States military could have mitigated the genocide in Rwanda, becomes a specific question about how quickly different configurations of US forces could have arrived, and an outcome of how many days it would take for a given force to arrive.[42] This is a familiar process for social science researchers, who are accustomed to defending their selection of dependent and proxy variables to answer the larger questions motivating their research. Researchers must explain why the outcome chosen is indeed a critical factor in the larger campaign: Kuperman had to first defend his argument that the time it would have taken US forces to arrive in Rwanda would have been the key to mitigating the genocide, in order to justify a campaign analysis focused on modeling time to arrival. Seemingly obvious "measures of effectiveness" (e.g. how many German planes were shot down by merchant ships' anti-aircraft guns) can obscure better measures of effectiveness (e.g. how many Allied ships survived crossing the Atlantic).[43]

Campaign analysis outcomes can be framed as either precise outcomes or sufficiency outcomes. For example, a researcher may focus on predicting the precise number of mines that Iranian forces could lay in the Strait of Hormuz undetected, or could ask the simpler question of whether Iran could lay at least one minefield.[44] Sufficiency outcomes often provide the answers that scholars want, and are often much easier to model in the context of great uncertainty than precise point estimate outcomes because more variables can be safely omitted (see Step 3). Researchers may, on the other hand, seek to estimate the precise number of nuclear warheads that could survive a counterforce attempt, or the number of casualties one state should expect to suffer in an invasion. The objective of the researcher– sufficiency or precise estimates– has implications for the appropriate treatment of uncertainty.

Step 2: Scenario

A key component of campaign analysis, often done in tandem with question selection, is defining the scenario, or the political-military context within which the interaction of military forces occurs. Defining the scenario for a campaign analysis involves making explicit choices about how to incorporate the political backdrop into the analysis of military operations. Scenario development requires researchers to identify the political factors that would most directly and determinatively shape the interaction of military forces under analysis, decide whether to hold these political variables constant by building them into the scenario, or to vary them to explore how military operations might unfold under different political conditions.

Scenario development begins with identifying the political variables most relevant to the interaction of military forces. These political variables may include whether the fight comes out-of-the-blue or in a moment of escalating crisis, the stakes of the conflict for each side, whether an alliance would be likely to hold or crack under pressure, all of which affect the disposition of forces and the scope of the conflict. Researchers employ a range of research methods to identify the political variables that most decisively shape the military operations

[42] Kuperman, *The Limits of Humanitarian Intervention*, 66f.

[43] Morse and Kimball, *Methods of Operations Research*, 52–53.

[44] Talmadge, "Closing Time."

of interest. They may draw on their knowledge of international security theory or their area knowledge, analyze past conflicts, or consult experts. As in any research design, researchers must be prepared to defend their choices of what political variables warrant discussion.

Researchers may also directly investigate how different political conditions might affect military outcomes, if this has bearing on their motivating question. Most researchers employing campaign analysis have chosen to hold political conditions constant in order to analyze military operations under specified political conditions. Recent work by Wu, however, explicitly varies one critical political variable, whether the nuclear counterforce attempt comes in peacetime or in crisis, to see how the military outcome would vary across different political contexts.[45] While researchers have not often employed campaign analysis to examine how changes in political variables might shape military outcomes, wargames and tabletop exercises are often explicitly designed to do just that, and methodological cross-pollination could expand the range of questions campaign analysis could be used to address.

Researchers choosing to set a single political backdrop for analysis of military operations have employed three general approaches to scenario development. A "most plausible" approach to scenario development t aims to identify the most likely political context that could give rise to a military operation of interest. Researchers using this approach employ their substantive knowledge and conduct research to determine the political pathways that might spark the military engagement, the resources each side would likely commit, and the possible involvement of allies. A "conservative" approach to scenario development is appropriate for researchers seeking to make a sufficiency argument. If a researcher is arguing that success in an operation is highly likely or highly unlikely, defining a "hard" scenario can make their conclusion more robust. For example, Bell strengthens his claim that British forces could defend the Falkland Islands by examining the campaign in a worst case scenario for the British: a no-warning attack.[46] Researchers have also selected "high leverage" scenarios, which may be implausible or an "easy test" for their conclusion, but they do so anyway because showing that an operation is possible can have important implications. For instance, a nuclear first strike with no prior warning is extremely unlikely, but a successful counterforce attack under these conditions would be an important indication of some degree of nuclear vulnerability.[47] High leverage scenarios might serve the public debate by demonstrating the political-military conditions within which Russia's secure second strike can no longer be assured. Moreover, when it comes to war, nations routinely plan for low probability, high impact possibilities, and their planning decisions can have important implications.

Researchers must take care to avoid a mismatch between the scenario and their claims. For example, if the researcher aims to claim that NATO forces would likely be able to withstand a Warsaw Pact invasion, then examining military operations in a political context favorable to NATO would not justify the broader claim.[48]

---

[45] Wu, "Living with Uncertainty.", pp. 86f.

[46] Mark S Bell, "Can Britain Defend the Falklands?" *Defence Studies*, Vol. 12, No. 2 (2012), pp. 283–301, doi:10.1080/14702436.2012.699726.

[47] Lieber and Press, "The End of MAD?"

[48] Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe."

Step 3: Model

Campaign analysis models, stated mathematically or in words, identify how variables interact in a scenario so that researchers can answer specific questions about important military operations despite the immense complexity and uncertainty of combat. The model should ideally be transparent, formalized, and explicit so that readers can easily understand how the outcome is being generated. As a report on military modeling put it, "A model is a mathematical or otherwise logically rigorous representation of a system or a system's behavior.''[49] The model consists of explanatory variables and the outcome they combine to produce.

Most published campaign analyses do not present their models in mathematical form, but almost all of them can be easily written down formally. Making the model formal(izable) has several important benefits. First, having a model that can be written down formally aids the researcher, clearly splitting the problem into individual variables that can be studied and estimated. Second, clear models help other researchers to debate the model itself, including whether the model contains the right variables and combines them appropriately. Finally, a formal(izable) model allows interested researchers to examine how an outcome would change with different input values and to adapt the model to other situations. We discuss this point further in our section on proposed recommendations.

As Talmadge explains, campaign analyses with transparent models can

> encourage rigor in the public debates that inevitably occur, by showing how different assumptions and data about military capabilities generate different predictions about the parameters of potential conflict [...] Analysts may still disagree, but at least they and those listening to them can ascertain the basis of their differences.[50]

Model validation is difficult because war is (thankfully) rare, and standard validation techniques in the modeling and simulation literature depend on observing outcomes across many separate observations.[51] Researchers should do what they can, however, to validate their models. The literature on modeling and simulation for operations research has some guidance on model validation that is useful in campaign analysis.[52] In some cases, researchers will be able to triangulate their findings between two different models. If both return the same result, their confidence in each model is strengthened. O'Hanlon uses this validation technique by studying a North Korean invasion using both a theater-level model and a zoomed in model of a specific battle, reaching similarly pessimistic conclusions about North Korea's prospects.[53] The simulations literature offers some useful techniques for locating coding errors in larger campaign analysis models: researchers can check that the model handles extreme values correctly and that outputs move in the expected direction as

---

[49] Paul K Davis and Donald Blumenthal, "The Base of Sand Problem: A White Paper on the State of Military Combat Modeling" (Arlington, VA: Defense Advanced Research Projects Agency, 1991), p. 1

[50] Talmadge, "Closing Time.", p. 84.

[51] James S Hodges, James A Dewar, and others, *Is It You or Your Model Talking?: A Framework for Model Validation* (Santa Monica: RAND Corporation, 1992). In at least one case, though, a hypothetical campaign actually later occurred: Shifrinson and Priebe, "A Crude Threat."

[52] Robert G Sargent, "Verification and Validation of Simulation Models," in, *Proceedings of the 2010 Winter Simulation Conference* (IEEE, 2010), pp. 166–183, doi:10.1109/WSC.2010.5679166

[53] Michael O'Hanlon, "Stopping a North Korean Invasion: Why Defending South Korea Is Easier Than the Pentagon Thinks," *International Security*, Vol. 22, No. 4 (1998), pp. 135–170, doi:10.1162/isec.22.4.135.

inputs change.[54] Researchers may also employ the "face validation" technique, in which people who have planned similar operations or who have knowledge of similar historical operations can examine the model and flag any important excluded variables.[55] The limitations on model validation make the process of model construction described below all the more important for the campaign analysis researcher.

The form that the model takes should be driven by the specific research question motivating the study, rather than by the desire to reflect reality in the finest detail possible. Not every detail of a fight will be relevant to the question the researcher is interested in answering, and efforts on the part of the researcher to include every possible variable to achieve maximum realism will at best have diminishing returns and at worst undermine the researcher's efforts to answer the question. In his review of Defense Department modeling efforts, Davis points out that "reality" on its own is a poor criterion: "Field exercises have real mud, noise, and confusion, but some runs of a computer simulation may have more realistic scenarios and better predictions of enemy tactics. Moreover, simulator training can be far more "real" in learning how to deal with extreme circumstances than field exercises with safety and environmental constraints."[56]

In general, the model should be the simplest model that provides "useful" results for the question being asked.[57] Simple models have four advantages. First, more complicated models are more resource intensive.[58] It is often easier to assign plausible value ranges to high-level aggregate variables (e.g. average convoy speed) than to sub-variables (e.g. speed of a specific type of tank across on a specific type of surface with stoppage time estimates based on refueling efficiency). A researcher might simply examine historical road marches to determine upper and lower bound or average convoy speed estimates rather than develop an extremely detailed model of every source of road march delay.

Adding complexity to models increases the opportunities for error in implementation. An early review of the TACWAR military modeling program found over 70 programming errors, ranging from failures to reset counters, to division by zero errors, to errors in how aircraft are allocated to close air support or attacks on enemy air bases.[59] As model complexity grows, it becomes more difficult to ensure the model is performing well and to diagnose the source of model errors. As a result, increasing the complexity of a model can reduce accuracy because of errors in the model.[60]

More complicated models are also more difficult to interpret. As models add variables and complex relationships between variables, it becomes more difficult for the modelers themselves to understand why the model may be producing a particular outcome, to check that its assumptions are met, or to communicate the model's results clearly to readers.[61]

---

[54]Sargent, "Verification and Validation of Simulation Models."

[55]Sargent.

[56]Davis, "Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation.", p. 1139

[57]Smith, "Essential Techniques for Military Modeling and Simulation.". This statement will be familiar to most social scientists, who are often issued with a George Box quote in their first semester of graduate school about "all models are wrong, but some are useful." George EP Box, "Science and Statistics," *Journal of the American Statistical Association*, Vol. 71, No. 356 (1976), pp. 791–799, doi:10.2307/2286841.

[58]Stewart Robinson, "Tutorial: Choosing What to Model–Conceptual Modeling for Simulation," in, *Proceedings of the 2012 Winter Simulation Conference (WSC)* (IEEE, 2012), pp. 1–12, doi:10.1109/WSC.2012.6465308.

[59]John C Ingram, "A Detailed Review of the TACWAR Model" (Harry Diamond Labs, Adelphi, MD, 1980).

[60]Robinson, "Tutorial," 1916.

[61]See O'Hanlon, "Stopping a North Korean Invasion.", p. 154.

Having models that can be easily understood and checked by readers should increase their ability to detect problems in the model and evaluate it, and therefore the confidence they can have in the results.

Given the costs of model complexity as well as the costs of omitting important variables, researchers conducting campaign analysis must make careful decisions about which variables to include and which to exclude from their models.

There are several methodologically sound reasons to exclude variables from the model. The first is that the variable is entirely irrelevant to the outcome of the military operation in question. For instance, a campaign analysis focused on Iran's ability to strike Saudi Arabia's oil infrastructure does not require any analysis of Iranian armored warfare capacity, because armored warfare has nothing to do with the fight.[62]

Second, the researcher may decide to exclude variables from the model that would affect the outcome modeled, but so marginally that the added precision is not worth the additional effort and model complexity. Well-trained crews executing good refueling operations can shave minutes of time off of each stop of a convoy. If the time to destination, however, depends largely on prompt mobilization, then any minutes a crew might be able to shave off on the road are swamped by the days it might take the nation (or coalition) to mobilize. This is especially true in the case of sufficiency outcomes, when the researcher can leave out variables whose cumulative effects would not be enough to change the sufficiency conclusion.

Third, a researcher's use of aggregated variables means that the constituent variables should be excluded. For instance, a nuclear model could use a warhead's "single shot probability of kill" against a target, meaning that the model should not then also include the (now redundant) components of SSPK: accuracy, yield, and target hardness.[63]

A final methodologically sound reason to exclude variables from the model is that a variable's inclusion would only strengthen the conclusion of a "sufficiency" argument. Mearsheimer, for instance, argued that NATO would be more competitive with the Warsaw Pact than the conventional wisdom of the day believed. He omits airpower variables from his model not because he thinks these variables are irrelevant to the outcome, but because he argues that airpower would favor NATO,[64] only strengthening his argument.[65]

There are unsound reasons to exclude model variables as well. Washburn and Kress, for instance, warn of an "ostrich" effect, where modelers might exclude a variable from the model simply because it is too difficult to set a value for it.[66] If a researcher discovers a critical variable cannot be estimated with any confidence, the researcher should either choose to evaluate how different values of that variable might shape the outcome of the analysis, or abandon the effort and ask a different question, rather than attempt to model the outcome without the variable and claim confidence in the result. As with all research designs, campaign analysis researchers must be transparent about their modeling choices, and ready to

---

[62] Shifrinson and Priebe, "A Crude Threat."

[63] Davis and Schilling, "All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask."

[64] Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe," 4, fn. 5.

[65] This reason depends on a broad agreement among readers in the direction of the effect. Cohen challenged Mearsheimer's assessment of airpower as NATO-favorable.

[66] Washburn and Kress, *Combat Modeling*.

defend their omissions. But also, critics should be prepared to explain why the variable that the researcher omitted, if included, would change the results of the analysis in a meaningful way. Modeling choices should be made and evaluated in their relationship to the question and argument.[67]

In addition to identifying the variables for inclusion in the model, the researchers must also assemble the variables into an equation (formalized or in prose). Researchers can draw on several techniques for assembling variables into an equation. Often, the model is a simple logical construction. If a researcher wonders whether a military force could execute a *fait accompli* before the other side could mobilize to defend the target, the researcher will need to estimate the time it will take for the first actor to send enough troops to seize the target. The time t needed for a convoy to move distance d given average speed v is simply $t = \frac{d}{v}$. A very simple model could consist of a set of "war stoppers," where a lack of any single variable would ensure failure.[68] Similarly, by the rules of probability, the probability that a target survives multiple independent attacks is the probability that it survives each of them, multiplied together.

Other models draw on existing models from physics, operations research, and military science to inform elements of model construction.[69] For instance, the physics of ballistics missiles and nuclear destruction are well summarized in the open source literature.[70] Lieber and Press are interested in the effects that increasing warhead accuracy and missile reliability have on US nuclear counterforce capability, so they must include weapon accuracy and reliability in their model. In their case, they can use published equations for the lethal radius LR in nautical miles of a nuclear weapon with a yield of Y in megatons and a target hardness H in pounds per square inch.[71]

Researchers often construct models of hypothetical military operations by examining historical operations. Raas and Long, for instance, construct a model of a hypothetical Israeli strike on Iranian nuclear facilities by examining the variables that mattered in the actual Israeli strike on Osirak in 1981.[72] Combining "workhorse" models and insights from historical operations, many researchers conducting campaign analysis have employed the controversial "3:1" rule, a very simple model derived from historical experience that posits that attackers with forces that are more than three times as great as the defender will usually pre-

[67] Criticisms of previous researchers' modelling choices can serve as the foundation for a new campaign analysis. See, e.g. Epstein, "Dynamic Analysis and the Conventional Balance in Europe."; Daryl G Press, "The Myth of Air Power in the Persian Gulf War and the Future of Warfare," *International Security*, Vol. 26, No. 2 (2001), pp. 5–44, doi:10.1162/016228801753191123.

[68] Kim R Holmes, "Measuring the Conventional Balance in Europe," *International Security*, Vol. 12, No. 4 (1988), p. 166, doi:10.2307/2539000.

[69] We encourage researchers to consult the military operations research literature as a starting point for model construction. See the journals *Operations Research* and *Military Operations Research* and the proceedings of the *Winter Simulation Conference*. For useful advice on when a model requires adaptation to a new question, see Richard J Hillestad, Bart Bennett, and Louis Moore, "Modeling for Campaign Analysis: Lessons for the Next Generation of Models. Executive Summary." (Santa Monica: RAND Corporation, 1996), pp. 13f

[70] Samuel Glasstone, *The Effects of Nuclear Weapons* (US Atomic Energy Commission, 1962); Davis and Schilling, "All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask."

[71] Specifically, LR $= \frac{1.45Y^{1/3}}{H^{1/3}} \left\{ \sqrt{1 + \frac{2.79}{H} + \frac{1.67}{H^{1/2}}} \right\}^{2/3}$. Glasstone, *The Effects of Nuclear Weapons*; Davis and Schilling, "All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask.", p. 213.

[72] Raas and Long, "Osirak Redux?"

vail.[73]

Researchers constructing models of *historical* military operations have additional options. McCue's book on antisubmarine warfare in the Bay of Biscay during World War II is a useful illustration of how researchers can construct models when they are examining campaigns that actually took place and have access to historical data.[74] McCue extends Morse and Kimball's wartime model of U-Boats in the Bay to include greater detail, using historical data that was available after the war, such as Dönitz's wartime diaries, to develop his more sophisticated model. Specifically, he models the repair capacity of shipyard in France, U-Boat circulation through the Bay, and attacks on convoys as one complete model, with more detail in some of the components. He reaches several interesting conclusions, including that Germany would have been much more effective if it had increased the number of at-sea resupply submarines. Regardless of how researchers construct their models, they will always face decisions about the appropriate level of resolution for modeling their campaign.

Step 4: Assign Values

Once the model is specified, researchers must assign values to each parameter in order to answer the question that motivates the study. Uncertainty in parameter values is a critical challenge for campaign analysis, and value assignment is therefore a crucial step for valid inference.

There are several ideal-type options for assigning values to model parameters given uncertainty. The first option is the "most plausible" approach, in which the researcher selects a "best estimate" parameter value. This approach is appropriate for parameters for which there is good enough data in the public domain to be confident in an estimate, which is often the case for "bean counts" of publicly known forces or other known quantities like the range of a particular type of helicopter. Very good estimates for many variables can be hard to find, however, and precise point estimates can be difficult to defend. It is especially difficult to assign values to model parameters for which information is classified or concealed.

Crucially, however, precise estimates of parameter values are often unnecessary, depending on the question posed by the researcher. An alternative strategy is to select "conservative" values with respect to the sufficiency outcome being estimated. If a value can be defended as an upper or lower bound on the variable and the analysis still reaches the same conclusion about sufficiency, readers can be confident that a more "accurate" value would not change the result. As Epstein puts it:

> Is that the 'right' Soviet value? Probably not. But is it unfavorable to the Soviets to use that value? Not in my judgment. And if, on assumptions of that sort, the Soviet still fail to execute the attack, then surely, on more "realistic" assumptions, they would fall even shorter of the mark.[75]

Similarly, Bell defends his conclusion that Britain could defend the Falklands by selecting

---

[73] Mearsheimer, "Numbers, Strategy, and the European Balance.", p. 176. For criticism, see Epstein, "Dynamic Analysis and the Conventional Balance in Europe."; John J Mearsheimer, "Assessing the Conventional Balance: The 3:1 Rule and Its Critics," *International Security*, Vol. 13, No. 4 (1989), pp. 54–89, doi:10.2307/2538780

[74] McCue, *U-Boats in the Bay of Biscay*.

[75] Epstein, *Measuring Military Power*, 199:xxviii.

values favorable to Argentina.[76] For instance, he posits that British reinforcements would begin arriving within 36 hours of an Argentine attack, a conservative assessment against the British given that expert opinion assessed that they would arrive within 24 hours. Uncertainty regarding a parameter value is thus not always a weakness of the analysis, but can be a source of strength. If a conclusion holds even in a worst-case scenario, the conclusion will stand for any plausible input. This is analogous to showing that a theory holds in a hard test case: "if the argument holds here, where the deck is stacked against it, it should hold elsewhere more easily."[77]

Practically, researchers can consult a number of sources in order to identify most plausible parameter values, most conservative values, and upper and lower parameter limits. Each data source is imperfect, and thorough researchers should triangulate parameter values against multiple data sources whenever possible.

Information for the initial "bean counting" of a campaign analysis can come from reported technical information on the number and performance of equipment and units involved. This information is often available from publications like Jane's, the IISS *Military Balance*, or think tank reports. Data on the order of battle or locations of forces are increasingly available using open source intelligence techniques.[78] Researchers may also look to similar historical military operations as a source of information for parameter values, especially for less "countable" or public parameters in the model.[79] For instance, in studying a potential North Korean invasion of South Korea, O'Hanlon draws on an Army study of historical rates of advance for armored units.[80] Data on historical operations is sometimes available in academic military histories, military after action reports, or even Congressional testimony, which Kuperman uses to estimate the speed of US strategic airlift.[81] In the best case, a previous analysis will be available, say from RAND, that proposes and defends reasonable values for a variable. Researchers may also consult experts, including military officers who have executed, planned, or practiced campaigns similar to the one being modeled, for their assessments of plausible values.

The choice between using most plausible parameter values and most conservative values depends on the question and available data. For a researcher asking a question requiring a precise outcome, such as "How long would it take a convoy to arrive at a target area?" the values in the model need to be accurate in order for the final outcome to be accurate. However, the "most plausible" approach is often impractical because available data is often inadequate to justify a precise chosen value for every model parameter.

For researchers asking sufficiency questions, such as "would an Indian convoy beat a Pakistani convoy in a race to a particular target area in Pakistan?" the conservative value assignment approach can be very powerful. A researcher's claim that India would win the race

---

[76] Bell, "Can Britain Defend the Falklands?"

[77] Bell, p. 286

[78] For example, Shifrinson and Priebe use Google Earth satellite imagery to identify aim points for an Iranian missile attack. Eveleth uses Google Earth and media reports to locate Chinese nuclear forces. Shifrinson and Priebe, "A Crude Threat."; Decker Eveleth, "Mapping the People's Liberation Army Rocket Force," https://www.aboyandhis.blog/post/mapping-the-people-s-liberation-army-rocket-force

[79] See Jacob A Stockfisch, "Models, Data, and War: A Critique of the Study of Conventional Forces" (Santa Monica: RAND Corporation, 1975), p. vii.

[80] O'Hanlon, "Stopping a North Korean Invasion"; Robert L Helmbold, "A Compilation of Data on Rates of Advance in Land Combat Operations" (Bethesda, MD: Army Concepts Analysis Agency, 1990).

[81] Kuperman, *The Limits of Humanitarian Intervention*.

is strengthened by value assignments deliberately skewed to favor Pakistan. Researchers often use a mix of "plausible" and "conservative" values for their variables because many arguments would not hold up to the extreme hard test of conservative value assignment for every model parameter.

In addition to the most plausible and most conservative approaches to parameter value assignment, researchers may also choose not to assign a *single* parameter value. They may instead decide on a range of plausible values for the model parameter. Then, they have at least two options. Existing research often re-runs analysis multiple times with multiple model parameter values (sensitivity analysis). Researchers can also vary all the distributions at once to produce a probabilistic range of outcomes (the input distribution approach using Monte Carlo techniques that we discuss in the Recommendations section).

Step 5: Run Model and Conduct Sensitivity Analysis

After constructing the model and setting values, researchers then run the model, plugging values into model parameters to produce the estimate of the outcome of interest.

Because researchers are rarely certain about the value of every parameter, they often conduct sensitivity analysis to show how outcomes are affected by changes in key input variables. The appropriate approach to sensitivity analysis should be tied closely to how the researcher approached value assignment, and the question and argument of the researcher. If a sufficiency argument holds up to an all-conservative value assignment, there is no need for further sensitivity analysis, because the argument has already withstood the hardest test. It is also methodologically defensible not to vary a parameter value if that parameter is known with high confidence.[82] In practice and due to space constraints, researchers often do not conduct sensitivity analysis for variables that they believe have little effect on the final outcome, but this can leave the research open to critique.

In standard sensitivity analysis, researchers identify a range of plausible values for a small number of model parameters, often an "upper" and "lower" limit. They then run the analysis twice, seeing how the outcomes change as the parameter values change. For instance, Shifrinson and Priebe have imperfect information about the accuracy of the Iranian missiles they consider in their analysis of a (now less) hypothetical Iranian attack on Saudi oil processing infrastructure, so they see how their findings change with different levels of missile accuracy.[83] Other researchers identify several key parameters and run the model multiple times using different sets of parameters.[84]

Step 6: Interpret and Present Results

The final step in campaign analysis is interpreting the output of the model and presenting the answer to the motivating question. While the model may produce a numerical output, the answer to the motivating question will often be presented in words. Researchers must take care to present answers with appropriate uncertainty (as in all social science research). As Kupchan puts it, "Making explicit the full range of political and strategic assumptions

[82] "Our results are robust to France having two aircraft carriers instead of one."
[83] Shifrinson and Priebe, "A Crude Threat," 191.
[84] E.g., Posen, *Inadvertent Escalation*

that produce a given output does not obviate the need to improve confidence levels and to include error terms with all assessments."[85]

The way the outcome is presented will affect reader interpretation. Presenting the result as a yes/no finding ("the most likely outcome is that all nuclear weapons are destroyed") conveys different meaning to readers than a probabilistic statement such as "the probability that at least one nuclear weapon survives is 30%," although both could simultaneously be true. Researchers should also consider whether to state a probabilistic form in numerical terms ("the probability that at least one nuclear weapon survives is 30%") or in qualitative terms ("a nuclear weapon will probably not survive").[86]

Researchers should resist the temptation to over-extrapolate the results from one scenario to a broader conclusion. For instance, rather than claiming that NATO *would* win in Europe, Posen took care to note only that the common assessment of "NATO's weakness on the ground, was at least open to challenge."[87] Mearsheimer, in contrast, makes a broader claim from a similar scenario, arguing that "NATO's prospects for thwarting a Soviet offensive are actually quite good."[88] Cohen was correct to take issue with Mearsheimer's conclusion "that the conventional balance in Europe is adequate on the basis of a single scenario resting on highly questionable political premises."[89]

From an ethical perspective, researchers should be attentive to how the findings might be interpreted or used by decision-makers to justify different policies. For instance, far from advocating for the bolt-from-the-blue counterforce strike their analysis suggests the US would be *capable* of executing, Lieber and Press warn that the pacifying effects of the nuclear revolution may be undone, and suggest US decision-makers carefully consider how further counterforce improvements might affect nuclear stability.

Two Recommendations for Researchers

We provide two recommendations for advancing the method of campaign analysis. Both recommendations would help researchers make better use of the qualitative research they have already conducted and to better manage uncertainty. The first recommendation is to treat the model as its own contribution and make it transparent and easily usable by other researchers. Second, we propose that researchers treat some or all input values as distributions, rather than single point estimates. The "input distribution approach" propagates input uncertainty through to the final output using Monte Carlo methods, strengthening inference despite input uncertainty with little additional effort.

[85] Kupchan, "Setting Conventional Force Requirements.", p. 572

[86] Scholars and practitioners of intelligence, beginning with Sherman Kent, have debated whether probabilities should be expressed in words or numbers. Sherman Kent, "Words of Estimative Probability," *Studies in Intelligence*, Vol. 8, No. 4 (1964), pp. 49–65. More recent literature suggests that expressing numerical probabilities improves forecasting accuracy, readers' comprehension, and may reduce overconfidence: Jeffrey A Friedman, Jennifer S Lerner, and Richard Zeckhauser, "Behavioral Consequences of Probabilistic Precision: Experimental Evidence from National Security Professionals," *International Organization*, Vol. 71, No. 4 (2017), pp. 803–826, doi:10.1017/S0020818317000352; Jeffrey A Friedman et al., "The Value of Precision in Probability Assessment: Evidence from a Large-Scale Geopolitical Forecasting Tournament," *International Studies Quarterly*, Vol. 62, No. 2 (2018), pp. 410–422, doi:10.1093/isq/sqx078

[87] Posen, *Inadvertent Escalation*, 68.

[88] Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe.", p. 3

[89] Cohen, "Toward Better Net Assessment.", p. 57.

Recommendation 1: Publish a Transparent Model

One of the intellectual contributions of a campaign analysis is the model developed in the analysis. The model should be presented in a clear way that helps other researchers evaluate and replicate it, and, when possible, researchers should make their code or spreadsheets available to help other researchers employ their models. Presenting the model in a clear and transparent way offers several benefits.

First, clear presentation of the model allows for better sensitivity analysis, both by the researcher and by potential critics or users. Readers with different information or different beliefs about a parameter value can quickly assess how conclusions change as inputs change if they have access to the model. This allows readers to conduct sensitivity analysis as well as the original researchers. Publishing the model in an easily reused form helps update campaign analyses as technology or weapons change. For instance, Shifrinson and Priebe build a model of an Iranian missile strike on Saudi oil infrastructure, flagging weapon accuracy as the key determinant of Iranian success.[90] They published the math behind their model, allowing researchers skeptical of any of their variable estimates to re-implement the model in code to see if their conclusion stood. In September 2019, eight years after Shifrinson and Priebe published their analysis, the availability of accurate cruise missiles enabled the successful attack on the Abqaiq oil processing facility in September 2019, just as their 2011 model suggested.

Second, the clear presentation of the model enables future researchers to answer new research questions or study new scenarios. The researcher who developed a model might be focused on a specific variable, leaving room for future researchers who might be interested in a different variable within the same scenario. Models developed for a specific question and scenario can often serve as a foundation when studying similar types of engagements. Some of the models used by researchers may become "workhorse" models, such as Kugler's FEBA expansion-attrition model used by Posen and O'Hanlon[91] or Epstein's logistics model serving as the foundation for work by Baker.[92] The values of variables within the model will always change across contexts (different countries have different numbers of warheads), and the model itself may need modification (suppression of air defense models should now account for cyber effects), but published models can serve as starting points for future researchers answering related questions in new contexts.

Recommendation 2: An Input Distribution Approach

Uncertainty in parameter values is a critical challenge for valid inference in campaign analysis. Researchers often work hard to understand the range of plausible values for each variable. In the existing approach to campaign analysis, researchers plug in a single value for each variable (or a small number if they conduct sensitivity analysis), producing single, point estimate outcomes from the model. Researchers often have much more uncertainty about inputs than a point estimate conveys, based on the substantive research they conduct. We propose an approach to addressing uncertainty and conducting sensitivity analysis that

[90] Shifrinson and Priebe, "A Crude Threat."

[91] Posen, "Measuring the European Conventional Balance"; Posen, *Inadvertent Escalation*; O'Hanlon, "Stopping a North Korean Invasion."

[92] Joshua M Epstein, *Strategy and Force Planning: The Case of the Persian Gulf* (The Brookings Institution, Washington, DC, 1987); Baker, "Logistics and Military Power."

propagates the knowledge researchers have about the field of variation around their inputs into their model to produce a probabilistic range of outcomes.

In this "input distribution" approach, researchers do not attempt to defend a single value for each variable, but instead quantify their uncertainty about each variable in the form of a statistical distribution. The output of the model itself becomes a distribution, reflecting uncertainty in the input variables. Instead of running the analysis several times with different values to produce several different outcomes, the input distribution approach enables researchers to conduct sensitivity analysis on all model variables at once and to present the outcome as a single distribution of outcomes.[93] This allows researchers to calculate likely values, confidence intervals, or other statistics to convey the uncertainty of the model's output.[94]

Modeling inputs as distributions cannot be done by hand, and requires computational tools. Monte Carlo techniques provide a simple way to implement the approach. A Monte Carlo technique consists of defining the distribution of all model variables, repeatedly sampling values from these distributions, and plugging each draw of variables into the model in order to produce a distribution of outcomes.[95]

As an illustration, a campaign analysis might hinge on the lethal radius of a nuclear warhead, modeled as a function of the warhead's yield and the target's hardness (see equation the lethal radius equation above 71). A traditional analysis could use a most plausible or conservative guess for yield and hardness and report two possible outcomes. An approach using statistical distributions to propagate uncertainty throughout the analysis would first specify distributions for yield and hardness (perhaps normal distributions, with means and variances). The Monte Carlo approach would then repeatedly draw values from each of the distributions, recalculate the equation, and generate a distribution over the lethal radius.[96]

The greatest benefit of using the input distribution approach is that it enables better sensitivity analysis than common approaches. Most importantly, it can capture interaction effects that would not appear when doing sensitivity analysis on a single variable at a time and

[93]Previous work has conducted sensitivity analysis on two variables at once, showing how the outcome changes as a function of both. Lieber and Press, "The End of MAD?".

[94]Ben-Haim suggests a method that flips the question: "how much error can we tolerate in making a decision?", studying the WEI/WUV algorithm that Posen draws on. Yakov Ben-Haim, "WEI/Wuv for Assessing Force Effectiveness: Managing Uncertainty with Info-Gap Theory," *Military Operations Research*, Vol. 23, No. 4 (2018), pp. 37–50, doi:10.5711/1082598323437

[95]As with many of the methods discussed in this paper, Monte Carlo methods were developed during WWII. John von Neumann, Stanislaw Ulam, Nick Metropolis, and others developed the technique to efficiently solve problems related to nuclear weapons design. Herbert L Anderson, "Metropolis, Monte Carlo, and the MANIAC," *Los Alamos Science*. There are rare circumstances where researchers could analytically compute the outcome distribution, rather than doing Monte Carlo methods. This will only be possible if all of their distributions are of a certain type, such as binomial or normal. Some previous campaign analyses have used Monte Carlo techniques, although in a quite different way than we propose. They use Monte Carlo techniques to show how the outcome varies as a single input takes on different values. They do not vary many inputs at once, suggest distributions for the inputs, or use the technique for propagating uncertainty. Bell, "Can Britain Defend the Falklands?"; Wu, "Living with Uncertainty."

[96]There are many similarities between our approach to campaign analysis and Bayesian search techniques, especially those used to find lost vessels underwater, which require users to set prior distributions and then use Monte Carlo techniques to propagate uncertainty through to the posterior. See, as examples, accounts of the search for the USS *Scorpion*, lost in 1968, and the SS *Central America*, lost in 1857. Richardson and Stone, "Operations Analysis During the Underwater Search for Scorpion."; Lawrence D Stone, "Search for the SS Central America: Mathematical Treasure Hunting," *Interfaces*, Vol. 22, No. 1 (1992), pp. 32–54, doi:10.1287/inte.22.1.32.

that the researcher might have otherwise missed. Two important variables, both set at extreme values, can produce a more extreme result than a researcher would obtain by varying each separately. (This is indeed what happens in the blast damage effect when yield and hardness both vary at the same time). Researchers can also conduct sensitivity analysis on all variables at the same time, showing how their results would change or determining that their arguments are robust to the full range of possible input values.

An output distribution also provides more information about the results of the analysis. For precise outcomes, adding uncertainty to the output lets researchers report not only the mean outcome, but also the 95% range, max, min, or any other statistic that might be useful. For sufficiency outcomes, the approach can quantify the probability that forces are sufficient, given uncertainty about the inputs. In keeping with most statistical methods in social science, which are very concerned with appropriately measuring and reporting uncertainty, the input distribution approach allows the uncertainty that researchers have in their inputs to be more fully reflected in their outputs. Existing campaign analyses add uncertainty to the output of the model heuristically, by interpreting the model's output in the context of their knowledge of the case, qualifying the model's output using their assessment of the inputs' uncertainty. The input distribution approach enables them to put more of the uncertainty directly into the model's output, rather than adding it on afterward.

The input distribution approach is not a panacea, however. If researchers select the wrong ranges, variance, or distributions for their inputs, the input distribution approach will produce incorrect outcome distributions. Selecting statistical distributions for inputs is the least familiar step in this advancement, so we provide guidance for how to do so in online Appendix B.

The input distribution approach only accounts for uncertainty in inputs to the model, not uncertainty about the model itself. If a model is misspecified, the actual outcome could be far outside the distribution it returns.[97] Moreover, the output distribution produced by the input distribution approach is only as useful as the research that informed the distributions. Poorly specified models or values based on misinformed research will produce an output distribution that looks impressive but ultimately does not improve our understanding of the world.

Researchers might also be tempted to neglect the research they have done and *over*count uncertainty. Researchers should use their expertise and research to set plausible ranges, rather than set ranges so wide that they encode no substantive knowledge and produce confidence intervals so wide that the researcher is no better off than before the analysis began.

It is also important to understand that the input distribution approach may not help researchers in situations where their questions are already answered adequately with existing approaches to sensitivity analysis. This is especially true in cases where the researcher is making a sufficiency claim and uses all-conservative values to test it. If a sufficiency finding is robust to using worst-case values for every input, the input distribution approach will add no further confidence in that conclusion, which has already withstood the hardest test. Often, though, researchers use a mix of plausible and conservative values or would like to produce a plausible outcome rather than a sufficiency outcome. The input distribution

[97] This is a familiar problem in statistical social science: the confidence interval on a regression coefficient reflects only sampling error, not the possibility that the model is wrong.

approach is advisable for these researchers.

We propose the input distribution approach as a new option for the management of parameter uncertainty. Although the suitability of the input distribution approach will depend on the question the researcher asks, the argument the researcher makes, and the data available for analysis, the approach, if done well, is an improvement over existing approaches under most circumstances. As with the development of the model (Step 3) and the assignment of parameter values (Step 4), the usefulness of the input distribution approach will rest largely on the quality of the substantive research that informs it.

Campaign Analysis in Practice

In this section, we replicate and extend two campaign analyses, Wu Riqiang's 2020 analysis of the United States' capacity to eliminate China's nuclear arsenal in a counterforce strike, and Barry Posen's 1991 analysis of NATO's capacity to forestall a Warsaw Pact invasion.

We select these two campaign analyses for replication and extension for several reasons. First, we are *able* to replicate them because they are exemplary in how transparently they document their models and parameter values. Second, the two analyses help to illustrate how the six steps of campaign analysis apply across scenarios that span conventional and nuclear warfare, operational and campaign levels of warfare, different regions of the world, and a 30-year time period. Third, these campaigns illustrate the value of our proposed advancements of the method, a focus on transparent and reusable models and the input distribution approach. We apply a different nuclear counterforce model, by Lieber and Press, to the scenario studied by Wu and reach similar findings about the survivability of the Chinese nuclear arsenal, despite differences in the two models.[98] By doing so, we show the reusability of campaign analysis models and thus the broader contribution a single campaign analysis model can make to the field of security studies. We use our replication of Posen's analysis to demonstrate the value of the input distribution approach to uncertainty. Our approach to propagating uncertainty serves as a robustness check, strengthening Posen's overall findings while showing greater variability in possible outcomes. We make our replications of the campaign analyses conducted by Wu, Press and Lieber, and Posen available as interactive calculators online for other researchers to employ.[99]

US-China Nuclear Counterforce: Replicating Wu (2020)

How survivable is the Chinese nuclear arsenal? Contemporary Chinese nuclear forces present somewhat of a puzzle for nuclear theorists. Despite having the two largest nuclear powers as potential adversaries, the United States and Russia, China has maintained a comparably small arsenal. A series of articles have examined the nuclear escalation dynamics and survivability of Chinese nuclear forces, especially against a US attack.[100] Most

---

[98] Lieber and Press, "The End of MAD?"

[99] Our tools are available at https://campaign-analysis.shinyapps.io/Inadvertent_Escalation/ , https://campaign-analysis.shinyapps.io/Nuclear_Counterforce/, and https://campaign-analysis.shinyapps.io/Merits_of_Uncertainty/

[100] Charles L Glaser and Steve Fetter, "Should the United States Reject MAD? Damage Limitation and US Nuclear Strategy Toward China," *International Security*, Vol. 41, No. 1 (2016), pp. 49–98, doi:10.1162/ISEC_a_00248; Talmadge, "Would China Go Nuclear?"; Fiona S Cunningham and M Taylor Fravel, "Dangerous Confidence? Chinese Views on Nuclear Escalation," *International Security*, Vol. 44, No. 2 (2019), pp. 61–109,

recently, Wu has used campaign analysis to argue that the Chinese nuclear deterrent was far from assured at several points in its past.[101]

In his article, Wu develops a nuclear counterforce model to examine the survivability of Chinese nuclear forces in eight different scenarios, facing attacks by either US or Soviet forces, in several years, under both peacetime and alert conditions. His principal conclusions are that China has retaliatory capacity in six of the eight scenarios, with the exception being a United States attack on Chinese nuclear forces in 2000, even when Chinese nuclear forces are on alert. He argues that 2010, when the introduction of road-mobile missiles vastly increased the probability of a warhead surviving, represents "a baseline for stable mutual deterrence."[102] By using one model and modifying model parameters to analyze different scenarios across two different dyads over a 25 year period, Wu's work is a rare exemplar of using a single model to answer questions about multiple scenarios (Recommendation 1).

Six Steps

Wu seeks to address the *question* of how Chinese nuclear survivability has evolved over time. The *outcome* Wu models to answer his question is the probability that the Chinese state could retaliate with a specified number of nuclear weapons after an attack on its nuclear forces by its principal adversaries under different scenarios over time.

Wu considers eight variations on a nuclear counterforce *scenario* to estimate the survivability of Chinese nuclear forces: a Soviet Union attack on Chinese nuclear forces in 1984 and US attacks on Chinese forces in 2000, 2010, and in an imagined 2025. For each of these possible scenarios, Wu examines both alert and non-alert conditions, resulting in a total of eight different scenarios.

Wu's *model* is well constructed to answer his specific question. Because he is interested in the probability that Chinese nuclear forces survive, he multiplies the probabilities that all the necessary components of retaliation survive an attack. For instance, for a mobile missile to successfully retaliate, it must survive an attack on its garrison, have prepared launch sites that have survived destruction, function properly when launched, and not be intercepted by ballistic missile defense. His model incorporates the probability of detection and destruction into each step, allowing for the possibility that the attacker does not have perfect information on the locations of all targets in China. Because China's nuclear arsenal is small, the model assumes that the attacker will use enough high accuracy/high yield warheads to reach a specified probability of destruction for each target. The model thus does not include a detailed treatment of weapons' accuracy, lethal radius, and target hardness,

---

doi:10.1162/isec_a_00359; Michael Chase and Evan Medeiros, "China's Evolving Nuclear Calculus: Modernization and Doctrinal Debate," in, *RAND/CNAC PLA Conference, Washington, Dc*; Sun Xiangli, "Analysis of China's Nuclear Strategy," *China Security*, Vol. 1, No. 1 (2005), p. 27; James C Mulvenon et al., *Chinese Responses to Us Military Transformation and Implications for the Department of Defense* (Rand Corporation, 2006); M Taylor Fravel and Evan S Medeiros, "China's Search for Assured Retaliation: The Evolution of Chinese Nuclear Strategy and Force Structure," *International Security*, Vol. 35, No. 2 (2010), pp. 48–87, doi:10.1162/ISEC_a_00016; Vipin Narang, *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*, vol. 143 (Princeton University Press, 2014); Fiona S Cunningham and M Taylor Fravel, "Assuring Assured Retaliation: China's Nuclear Posture and Us-China Strategic Stability," *International Security*, Vol. 40, No. 2 (2015), pp. 7–50, doi:10.1162/ISEC_a_00215; Riqiang Wu, "Certainty of Uncertainty: Nuclear Strategy with Chinese Characteristics," *Journal of Strategic Studies*, Vol. 36, No. 4 (2013), pp. 579–614, doi:10.1080/01402390.2013.772510.

[101] Wu, "Living with Uncertainty."

[102] Wu, p. 114

as other nuclear strike models do.[103]

Some *values for parameters* are easily estimated: the size of the Chinese nuclear arsenal and the warheads the United States would likely use are fairly well known. Other inputs, however, are known with much less certainty. Specifically, the probability that attacking forces could locate each target is very difficult to estimate. For *sensitivity analysis*, Wu shows how the probability of a warhead surviving varies across many values for the hardness of underground facilities, the alert rate of mobile missiles, and the effectiveness of United States ballistic missile defense. He does not conduct sensitivity analysis for some of the variables in his model, including the detection probabilities of different Chinese targets, which is a crucial variable for the analysis.

Wu *presents model results* in probabilistic terms. His model returns probabilities that different numbers of Chinese warheads are available for retaliation. For the US-China 2010 scenario, if the "criterion for deterrence" is a single warhead surviving to retaliate, he finds a 38% probability of meeting the single-warhead threshold when nuclear forces are on day-to-day alert, and 90% probability when the missiles are on fully alerted status.[104] The probability that 5 or more warheads survive is 6% and 1% for full alert and day-to-day alert, respectively.[105]

Studying a Chinese Counterforce Attack Using Lieber and Press' (2006) Model

We successfully replicate Wu's model of the US-China 2010 scenario and reach the same probabilistic conclusions. We choose to replicate the US-China 2010 scenario (one of eight scenarios Wu modeled), simply because Wu identified this as the "baseline for China-U.S. strategic stability,"[106] when mobile missiles first created a survivable deterrent.

We argue in Recommendation 1 that models are often an intellectual contribution in their own right. We demonstrate the broader applicability of models, and thus the wider contributions a single campaign analysis can make, by re-examining Wu's 2010 US-China counterforce attack using a slightly modified form of a previously published counterforce model by Press and Lieber.[107] Although the models are quite different, we find similar results using point estimates from Wu's paper in Lieber and Press' model. Our findings with respect to Chinese nuclear survivability when we use Lieber and Press' model are very similar to the findings when we use Wu's model, supporting our argument that models can be adapted for new questions, and that researchers make a major contribution beyond analysis of a single scenario when they publish their models.

The Lieber and Press model and the Wu model differ in several important respects. First, the two models have different outcomes. Lieber and Press' model estimates the number of warheads that are expected to survive a first strike, while Wu models the probability that they would survive *and* successfully strike the attacking country, accounting for missile reliability and ballistic missile defense. Next, Wu's model accounts for the possibility that missiles are dispersed and in locations not known to the attacker. To successfully retaliate, a

---

[103] Lieber and Press, "The End of MAD?"
[104] Wu, "Living with Uncertainty," A2.
[105] Wu, A2.
[106] Wu, 87.
[107] Lieber and Press, "The End of MAD?"

Chinese missile must survive at each step leading up to launch. Press and Lieber's model is built from a "bolt-from-the-blue" scenario, where all launchers are de-alerted and in known positions. The Lieber and Press model includes a detailed weapons effect model to estimate the probability that a target with a given hardness would be destroyed by a warhead with a given yield and accuracy. Wu's model black boxes this process, assuming a fixed probability of destruction for each target (100% for soft targets). Wu can safely exclude these details from his model because an attacking force would not be constrained in the number or type of warheads it could use, given the small size of the Chinese arsenal.

As part of our proposed emphasis on model transparency and reuse, we wanted to see if we could arrive at the same conclusions as Wu in the US-China scenario by adapting Lieber and Press' existing counterforce model. We begin with Lieber and Press' model of nuclear combat, which involves the specific accuracy and yield of different American launchers and warheads, weapon reliability for each kind of launcher, and the number and hardness of different target facilities. We needed to make two modifications to the Lieber and Press model to make it applicable to a Chinese scenario. First, we allow mobile missiles to be deployed rather than being fixed in garrisons, because missile mobility is key to Wu's scenario and was assumed away for Lieber and Press' bolt-from-the-blue scenario. Second, we needed to include a term for detection probabilities in the model. Lieber and Press did not need to include a detection probability term because they examined a no-warning scenario, but detection probability becomes critical when mobile missiles are deployed in high alert scenarios. We incorporated detection probability by adding a single, aggregate term to the model for the probability of an attacker locating a deployed mobile missile, as opposed to individual detection probabilities for launch pads, forward bases, and technical sites as Wu's detailed, China-specific model does. We make this expanded model available as open source software.

After making these changes, applying the model to the 2010 US-China context involved simply changing the number and types of targets and the weapons used in the attack from a US-Russia 2006 scenario to a US-China 2010 scenario, using the values given in Wu's paper.

Our result for US-China counterforce 2010 using Lieber and Press's 2006 US-Russia counterforce model was very similar to Wu's much more granular, custom model. Specifically, we run the modified Lieber and Press model twice with different values for our aggregate detection term. If an attacker has a 0.95 probability of locating a mobile missile, the probability of at least one warhead surviving is 70%. Using a more conservative detection probability of 0.70, the probability that a single warhead survives is over 90%, and a 50% probability that three or more survive.

Because the Lieber and Press model and the Wu model differ in the precise outcomes they model (Lieber and Press model the number of warheads expected to survive a first strike, while Wu models the probability that at least one warhead would survive and successfully strike the attacking country), model outputs cannot be easily compared side by side. Substantively, however, both models produce the same conclusion: they suggest that the Chinese arsenal is best described as "first strike uncertainty": few enough Chinese nuclear forces would survive the attack to assure a retaliatory strike. However, the attacker's ability to destroy all warheads with certainty is far from assured.

(Re)assessing the Conventional Balance in Europe: Replicating Posen 1991

In the 1980s, security studies researchers and policymakers debated the balance of conventional forces in Europe. Soviet numerical superiority led many to believe that a conventional defense of Western Europe would be impossible, forcing NATO to rely on tactical nuclear weapons to stop a Soviet armored invasion. In this debate, Posen conducted a series of campaign analyses and argued that NATO forces, if given appropriate credit for their superior training, equipment, unit sizes, and logistical support, were more competitive with the Warsaw Pact than the conventional wisdom believed.

The Six Steps

Posen is interested in understanding the balance of military power between the Warsaw Pact and NATO. Breaking off one concrete piece of that broad, net assessment-level topic, Posen chooses to consider the specific, campaign-level *question* of whether a conventional Soviet attack, composed mainly of armored divisions would have generated a breakthrough of NATO lines. He explains that he chose this question because the outcomes of many historical armored battles hinged on breakthrough of the enemy line. The specific *outcome* Posen estimates is the supply and demand of forces on each side and whether NATO faces a shortfall of forces, and he defends this outcome as an appropriate measure for NATO's ability to prevent a Warsaw Pact breakthrough. Posen is careful to specify the *scenario*: a conventional Soviet attack, composed mainly of armored divisions, advancing into Western Europe in the 1980s.

To *model* the conflict, Posen adopts a model first developed by Richard Kugler called "attrition-FEBA [forward edge of the battle area] expansion." This model takes in several parameters, the most important being the rate of advance and attrition of the attacking force, the exchange ratio of losses between the forces, and the width of the front that each unit can hold. If the force required by the attacker, as estimated by the model, exceeds the available force available to the attacker, the attacker pauses its advance. Conversely, if the defender experiences a shortfall, the attacker has achieved a breakthrough and presumably victory.

Much of Posen's work is concerned with proposing and defending most plausible *values* for each of the parameters in the model, drawing on historical analogues when possible and expert assessment elsewhere. He begins by reflecting the conventional wisdom, picking inputs that are consistent with the convictions of the NATO pessimists. He then *runs the model* with values that are favorable to the Warsaw Pact. Under these conditions NATO forces face major shortfalls. Posen conducts extensive qualitative research to assign what he assesses to be more plausible estimates for these same six variables that give NATO appropriate credit for its strengths, and runs the model again with the NATO-favorable values he defends.

Posen finds that under the Pact-favorable assessments, NATO forces face a shortfall. If, however, the values he defends at length are correct, NATO forces would have been sufficient to prevent a Warsaw Pact breakthrough. In *interpreting and presenting* the results of the analysis, though, Posen is careful not to claim more about the general status of United States and Russian forces than can be said with the single campaign analysis study. He does not, for instance, claim that his analysis proves NATO would have successfully forestalled a

Pact invasion, or that his analysis would have held in different political contexts than those he articulated. Instead, he says that the common assessment of "NATO's weakness on the ground, was at least open to challenge"[108] Using slightly stronger language elsewhere, he argues that "Under relatively conservative assumptions, NATO's forces appear adequate to prevent the Pact from making a clear armored breakthrough."[109]

Replicating and Extending Posen's Model

We replicate Posen's model and extend his analysis by employing the input distribution approach to aggregate and propagate the uncertainty in all model parameters through to the output.

We conduct our replication based on the details laid out in Chapter 3 and Appendix 3 to Posen (1991), which describe the formulas involved in calculating the FEBA-expansion attrition model and the values used. The book is remarkable in its transparency: the models, parameters, and code are all reported. We build our own interactive model and are able to successfully reproduce Posen's results.

Whereas Posen conducts his analysis twice (once using Warsaw Pact-favorable point estimates and once using what his research suggests are more plausible NATO-favorable point estimates), we use the input distribution approach, incorporating the full range of values to produce a distribution of outcomes, showing the proportion of simulations in which NATO prevails. Using the Monte Carlo approach, we draw uniformly from these ranges and recalculate the outcome (the probability of facing a shortfall of forces) many times to determine the probability of NATO force shortfall (the output) from a large combination of inputs.

We use a uniform distribution because our intention is to determine whether Posen's conclusion is robust to equally weighting all of the values between the pessimistic conventional wisdom and the more NATO-favorable values Posen defends.

Figure 1?? shows the variance in several outcomes of interest that result from calculating the outcomes with random draws of the input variables from the full ranges, NATO-favorable and Pact-favorable, given in the book. Incorporating all of this uncertainty suggests that NATO forces are likely to hold out, but face a non-zero chance of being overrun. Toward the end of the 90-day campaign that Posen studies, the probability of a NATO shortfall occurring at any point approaches 25%, although Pact forces would also be depleted here. On the whole, however, NATO is competitive throughout the campaign under most combinations of variables. Examining other outcomes reveals a broad range of possible outcomes: the Pact can lose anywhere from 12 to 38 armored division equivalents, in contrast with the original estimate of 30, and NATO can lose between 4 and 22, in contrast to the original estimate of 5.

The principal difference between our results and Posen's results when we incorporate uncertainty directly is that Posen arrives at a binary conclusion that NATO would successfully forestall a Warsaw Pact invasion under the conditions he believes are accurate and would fail under what he considers implausibly pessimistic conditions. The uncertainty in Posen's findings comes from his substantive interpretation of the model's point estimate. In contrast, we directly incorporate Posen's substantive research about input uncertainty

[108] Posen, *Inadvertent Escalation*, 68.
[109] Posen, 127.

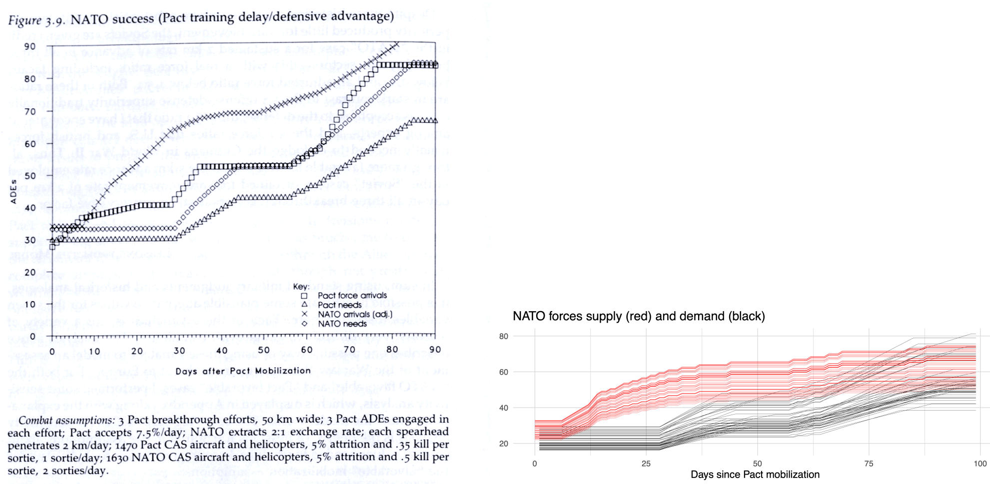

Figure 1: Figures showing Posen's original supply and demand curves for NATO and Pact forces (left) and NATO supply / demand curves using the input distribution approach (right). The y axis shows the number of armored division equivalents (ADE) and the x-axis shows the number of days since Pact mobilization. In Posen's NATO-favorable case, the line indicating NATO demand never passes above the line showing NATO supply, indicating that NATO does not face a shortfall. The variance in the ADEs demanded by NATO (black lines in the figure on the right) is caused by different Monte Carlo draws for how many ADEs are required to defend a length of front, and different draws for rates of losses caused by different rates of Soviet advance, exchange rates, and air power kills. In some of the draws (around 20%), demand exceeds supply for NATO forces, creating the potential for a Soviet breakthrough.

into an output that expresses NATO's competitiveness with the Warsaw Pactacross a range of possible conditions. Under some of these combinations of variables, NATO forces lose to the Warsaw Pact, but in most combinations of inputs, NATO forces are sufficient.

Overall, our analysis serves as a robustness check that ultimately supports Posen's substantive conclusion: NATO ground forces were likely more competitive with the Warsaw Pact than the conventional wisdom of the day suggested. Posen's overall assessments of the conventional balance, identification of most plausible pathways to nuclear escalation, and his proposed investments in NATO ground forces remain intact. Our analysis shows that Posen's conclusions stand even if we treat his well-researched most plausible values as simply the upper bound on a range of values that also include Pact-favorable inputs.

Contribution to Academic Theory:

Campaign analysis is a promising and underused method for measuring variables at the center of international relations theory. Enduring debates about the nuclear revolution and offense-defense theory can be advanced through more precise measurement of nuclear survivability and the ease of conquest. We show how campaign analyses like those conducted by Wu, Lieber and Press, and Posen can be used to improve on existing approaches to measuring these variables.

Alternative Measurement of the Nuclear Balance

The bedrock of nuclear deterrence and the pacifying influence of nuclear weapons in international relations theory is the secure second strike: the ability for a state to credibly threaten retaliation with surviving nuclear weapons after an initial attack.[110] Second strike survivability, according to nuclear revolution theory, means that states with secure second strike no longer need to fear conquest, and as a result, they can be less sensitive to the relative balance of power,[111] arms racing,[112] or competing for allies and territory.[113] In short, a large body of international relations scholarship considers nuclear arsenals pacifying to the extent that states are deterred from attacking each other by confidence in the target's ability to retaliate.

During much of the Cold War, precise measurement of secure second strike was not a central problem. The first nuclear era was characterized by large arsenals with limited counterforce capabilities, and from roughly 1963 onward practitioners and academics generally believed that both US and Soviet retaliatory capabilities were assured.[114] During this period, the

[110]See, for seminal examples, Jervis, *The Meaning of the Nuclear Revolution*; Kenneth N Waltz, "Nuclear Myths and Political Realities," *American Political Science Review*, Vol. 84, No. 3 (1990), pp. 730–745, doi:10.2307/1962764; Stephen Van Evera, *Causes of War: Power and the Roots of Conflict* (Cornell University Press, 1999)

[111]Jervis, *The Meaning of the Nuclear Revolution*; Waltz, "Nuclear Myths and Political Realities."

[112]Robert Jervis, "Why Nuclear Superiority Doesn't Matter," *Political Science Quarterly*, Vol. 94, No. 4 (1979), pp. 617–633, doi:10.2307/2149629; Waltz, "Nuclear Myths and Political Realities."

[113]Jervis, "Why Nuclear Superiority Doesn't Matter"; Van Evera, *Causes of War*; Robert Jervis, "Was the Cold War a Security Dilemma?" *Journal of Cold War Studies*, Vol. 3, No. 1 (2001), pp. 36–60, doi:10.1162/15203970151032146.

[114]Spurgeon M Keeny Jr and Wolfgang KH Panofsky, "Nuclear Weapons in the 1980s: MAD Vs. NUTS," *Foreign Affairs*, Vol. 60, No. 2 (1981), pp. 287–304; Harold Feiveson and Frank Von Hippel, "The Freeze and the Counterforce Race," *Physics Today*, Vol. 36, No. 1 (1983), p. 36, doi:10.1063/1.2915443; Robert Jervis, *The Illogic of American Nuclear Strategy* (Cornell University Press, 1984); Jervis, *The Meaning of the Nuclear Revolution*;Charles Glaser, "Why Do Strategists Disagree About the Requirements of Strategic Nuclear Deterrence?" *Nuclear Argu-*

secure second strike variable was implicitly treated by academics as a dichotomous variable that only took the value "exists." Because of general confidence among academics in the existence of secure second strike, there was little demand for a sophisticated methodology to more precisely measure the variable at the heart of the nuclear revolution.

However, a feature of the contemporary nuclear era is some loss of confidence in secure second strike due to smaller arsenals and improved counterforce, particularly improvements in missile accuracy and technical intelligence.[115] Because secure second strike can no longer be taken for granted, it is important to measure second strike security not on a dichotomous "yes/no" scale, but rather as a continuous variable representing the confidence both states might have that a target state might be able to inflict "unacceptable" retaliatory damage across a range of scenarios. Otherwise put, the key measurement question for several nuclear states has changed from "does a state have secure second strike or not?" to "how confident can a state be that it (or its target) will have some number of surviving warheads after an attack under specified conditions?"

Beyond the nuclear revolution, the nuclear survivability variable is also central to academic discussion of coercive diplomacy. Whereas Sechser and Furman argue that nuclear weapons provide little advantage in coercive diplomacy, Kroenig has argued that nuclear superiority creates a bargaining advantage in coercive diplomacy.[116] This debate is inconclusive, however, in part due to imprecise measurement of the central variable. Sechser and Furman code the nuclear balance as a simple binary: states either have nuclear weapons or they do not, and a nuclear balance can either be advantageous to one side or the other, or a "tie." Kroenig's approach to measurement of the nuclear balance is only a slight improvement. He counts the warheads on each side of a nuclear dyad to create a numerical ratio to represent the nuclear balance.

Campaign analysis presents a significant improvement in measurement that incorporates how intelligence for counterforce[117], command and control arrangements, and precision targeting ability combine under different political scenarios to determine target survivability. That is why Wu uses campaign analysis, rather than a cruder proxy, to measure the level of confidence China could have in its ability to strike back against Russian and US attacks under peace-time and alert conditions.

Alternative Measure of the Ease of Conquest

Although he did not frame his work this way, Posen's campaign analysis produced an improved measure of the ease of territorial conquest (under the specified conditions), a vari-

ments: Understanding the Strategic Nuclear Arms and Arms Control Debates, pp. 109–171. For an example of important work suggesting that Soviet Union leadership was not so sanguine about its nuclear survivability, and that US intelligence for counterforce was better than many believed, see Brendan R Green and Austin Long, "The MAD Who Wasn't There: Soviet Reactions to the Late Cold War Nuclear Balance," *Security Studies*, Vol. 26, No. 4 (2017), pp. 606–641, doi:10.1080/09636412.2017.1331639 and Austin Long and Brendan Rittenhouse Green, "Stalking the Secure Second Strike: Intelligence, Counterforce, and Nuclear Strategy," *Journal of Strategic Studies*, Vol. 38, Nos. 1-2 (2015), pp. 38–73, doi:10.1080/01402390.2014.958150, respectively.

[115] Lieber and Press, "The End of MAD?"; Keir A Lieber and Daryl G Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence," *International Security*, Vol. 41, No. 4 (2017), pp. 9–49, doi:10.1162/ISEC_a_00273.

[116] Todd S Sechser and Matthew Fuhrmann, *Nuclear Weapons and Coercive Diplomacy* (Cambridge University Press, 2017); Matthew Kroenig, *The Logic of American Nuclear Strategy: Why Strategic Superiority Matters* (Oxford University Press, 2018).

[117] Raas and Long, "Osirak Redux?", p. 31

able at the center of offense-defense theory. Offense-defense theory holds that war is more likely when conquest is easy, as it makes war initiation more profitable, encourages territorial expansion for defensive purposes, and creates incentives for states to strike first in a crisis.[118] The offense-defense balance, like the secure second strike, is a key variable in academic international relations.

The difficulties of measuring the offense-defense balance have long stymied academic debate around offense-defense theory.[119] The key military outcome at issue in offense-defense theory is whether the attacker can win an operation decisively enough to end a war quickly, and with relatively few casualties. When this can be done, offense is considered dominant and war is believed to be more likely.[120] Scholars have relied on a variety of measures that poorly approximate the offense defense balance, ranging from the technology across a dyad to "bean counts" of military forces. These measures are flawed because they fail to capture the ease with which one side would be able to accomplish a military objective vis-a-vis the other side. Other scholars provide a more complete definition of offense-defense balance, but not a clear approach for measuring it in practice.[121]

Biddle suggests an improved approach to measurement of the offense-defense balance focused on the adversaries' relative skill in force employment at the operational level of war.[122] Although a significant improvement over the bean count approach, Biddle's proposed force employment measure is abstract, unattached to the actual, scenario-specific operations that encourage or discourage aggression.

Glaser and Kaufmann suggest that campaign analysis can be used to improve measurement of the offense-defense balance. They note that "Measuring the offense-defense balance requires working through essentially the same steps as performing a net assessment: for given military missions, the analyst develops a model of how the forces will interact in combat, and explores the predictions of the model under different scenarios."[123] Although they use the term "net assessment," their description of the method is much closer to our definition of campaign analysis than to net assessment. In particular, they call on researchers to focus on the interaction of military forces in one "campaign or theater of operations whose outcome will be critical for the outcome of the entire conflict."[124] Specifically, Glaser and Kaufmann "calculate the offense-defense balance by comparing the cost of forces the attacker requires to launch a successful blitzkrieg to the cost of the defender's forces."[125] Otherwise put, Glaser and Kaufmann use campaign analysis to measure the offense-defense balance.

In agreement with Glaser and Kaufmann, we argue that campaign analysis improves upon alternative approaches to measuring the offense-defense balance by permitting researchers

---

[118] Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics*, Vol. 30, No. 2 (1978), pp. 167–214, doi:10.2307/2009958; Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?"; Van Evera, *Causes of War*.

[119] Jack S Levy, "The Offensive/Defensive Balance of Military Technology: A Theoretical and Historical Analysis," *International Studies Quarterly*, Vol. 28, No. 2 (1984), pp. 219–238, doi:10.2307/2600696; Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?"

[120] If the war would be long in duration or the costs high, defense is considered dominant and war is believed to be less likely.

[121] Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?"

[122] Stephen Biddle, "Rebuilding the Foundations of Offense-Defense Theory," *The Journal of Politics*, Vol. 63, No. 3 (2001), pp. 741–774, doi:10.1111/0022-3816.00086.

[123] Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?", p. 74

[124] Glaser and Kaufmann, 74.

[125] Glaser and Kaufmann, 52.

to model *scenario-specific* outcomes that are pertinent to the dyad in question. Rather than an overall assessment of technology in the international system–a crude proxy for the ease with which Russia could potentially storm the Suwalki Gap–campaign analysis equips the researcher with the tools to develop the most plausible possible estimate of the outcomes that offense-defense theory suggests drives or discourages aggression, such as the ratio of attacker casualties to defender casualties (the loss-exchange ratio), attacker casualties per square kilometer of territory conquered, or the probability of attacker victory *in a particular, plausible operation*. Posen, for instance, helped to improve measurement of the offense defense balance between NATO and the Soviet Union, which could have been leveraged by academics to defend a measure of defense dominance in the dyad during the period of study.[126] Today, campaign analysis could likewise be used to measure the offense defense balance between NATO and Russia, between North Korea and South Korea, between China and India, between India and Pakistan, or any other pair of potential adversaries, allowing researchers to test the predictions of the theory.

Conclusion

Military operations are central to international relations theory and practice, but until now the methodological guidance for academics analyzing military operations has been sparse. Recent work has formalized and advanced methods in wargaming and archival research, and we contribute to the growing methodological literature within security studies by defining and advancing the method of campaign analysis. Standardizing campaign analysis, as with other methods, equips readers to more easily evaluate its use, enables a wider pool of researchers to employ the method well, and creates a baseline from which future researchers may advance the method.

In addition to defining and standardizing the existing practice of campaign analysis, we offer two recommendations of our own. The first is an emphasis on the intellectual contribution that researchers make when they develop models for campaign analysis. Models can be applied beyond the specific scenario under investigation and can serve as the foundation for other researchers' models. Making models transparent and reusable would help further the wider community's research efforts and provide a lasting contribution beyond the specific scenario. As a gold standard, researchers could publish their models as interactive calculators (as we do in the replications) for other researchers to use and adapt. The second recommendation is a technique for propagating uncertainty through the campaign analysis. Much of the criticism of campaign analysis stems from disagreement about the precise values used in the study. Treating inputs as distributions rather than fixed points and propagating this uncertainty through to the final outcome using Monte Carlo techniques helps to address this concern.

This paper is motivated by the conviction that campaign analysis, carefully executed to manage uncertainty, equips scholars to contribute to the healthy function of the marketplace of ideas in defense policy and to advance theoretical debate. Campaign analysis studies can reveal theoretical puzzles, suggest new theories, and produce alternative measures for key

[126]Glaser and Kaufmann suggest that Posen's model could be used to measure their preferred definition of offense-defense balance, the ratio of attacking and defending forces above which the attackers prevail: "By re-running the model and increasing Warsaw Pact forces until they do prevail, we could estimate the balance for that scenario" Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?", pp. 75-76.

variables in theoretical debates. We demonstrate how campaign analysis can produce measures for two variables at the core of canonical international relations debates: nuclear survivability, and ease of conquest. Although these two variables are central to international relations debates around the nuclear revolution and the offense defense balance, the debates have largely relied on problematic proxy variables. Campaign analyses can provide alternative operationalization to advance debates stalled by imprecise measurement.

We see great potential for campaign analysis to answer an even wider set of questions. Campaign analysis could be fruitfully applied to more historical campaigns, helping us to understand why an operation turned out the way it did, validating models, and revealing puzzles. Researchers could also focus more often on examining the effects of key variables, rather than on the likely outcomes of operations. Substantively, a large set of security-related questions at the intersection of military studies and politics are amenable to campaign analysis but have not yet been addressed with the method. Most campaign analyses define a constant political landscape that limits the complexity of the model to facilitate focused attention to the interaction of military forces under specified political conditions, but there is no imperative to do so. Scholars might fruitfully employ campaign analysis to study how policies such as sanctions could shape military outcomes, or how alliance cohesion or fracture could shape conflict.

There is room for much more collaboration and cross-pollination between academic campaign analysis and related research methods employed in government and government-funded research centers. Academics could draw more from the existing models in the military modeling and sims literatures to construct the models in their own research, and they could borrow more from the operations research to answer optimization questions.[127] Models of decision-making, drawn from political science or using specific techniques such as wargames, table top exercises, and game theory, could help expand the range of questions and scenarios that researchers can study with campaign analysis.[128]

Ultimately, the value of campaign analysis depends on the researchers' motivating questions, substantive knowledge, and careful scholarship. Models will produce outputs, but outputs are only meaningful if researchers model outcomes that are relevant to their questions, identify the critical variables for inclusion in the model and defend the exclusion of others, conduct the research necessary to set reasonable parameter distributions, and carefully interpret their results. Researchers trained in the fundamentals of research design and equipped with substantive knowledge of international security are well positioned to employ campaign analysis to inform policy and advance academic debate.

[127]The 1954 RAND "basing report" is an example of a campaign analysis, by our definition, answering an optimality question. Albert Wohlstetter et al., "Selection and Use of Strategic Air Bases," *RAND Corporation*

[128]See, e.g. Hillestad, Bennett, and Moore, "Modeling for Campaign Analysis.", p. 21

Appendix A: Selected Bibliography of Campaign Analyses

Year	Author	Title	Publication
1954	Wohlstetter, Hoffman, Lutz, and Rowen	Selection and Use of Strategic Air Bases	RAND
1973	Davis and Shilling	All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask	*Journal of Conflict Resolution*
1976	Steinbruner and Garwin	Strategic Vulnerability: the Balance between Prudence and Paranoia	*International Security*
1981	Epstein	Soviet Vulnerabilities in Iran and the RDF Deterrent	*International Security*
1984	Posen	Measuring the European Conventional Balance: Coping with Complexity in Threat Assessment	*International Security*
1987	Epstein	*Strategy and Force Planning: The Case of the Persian Gulf*	Brookings
1988	Posen	Is NATO Decisively Outnumbered?	*International Security*
1988	Mearsheimer	Numbers, Strategy, and the European Balance	*International Security*
1988	Epstein	Dynamic Analysis and the Conventional Balance in Europe	*International Security*
1988	May, Bing, Steinbruner	Strategic Arsenals After START: the Implications of Deep Cuts	
1989	Mearsheimer	Assessing the Conventional Balance: The 3:1 rule and its Critics	*International Security*
1989	Salman, Sullivan, and Van Evera	Analysis or Propaganda? Measuring American Strategic Nuclear Capability, 1969-88	*Nuclear Arguments*
1990	McCue	*U-Boats in the Bay of Biscay: An Essay in Operations Analysis*	NDU Press
1994	Masaki	The Korean Question: Assessing the Military Balance	*Security Studies*
1996	Biddle	Victory Misunderstood: What the Gulf War Tells Us About the Future of Conflict	*International Security*
1998	O'Hanlon	Stopping a North Korean Invasion: Why Defending South Korea is Easier than the Pentagon Thinks	*International Security*
2001	Greenhill	Mission Impossible? Preventing Deadly Conflict in the African Great Lakes Region	*Security Studies*
2003	O'Hanlon	Estimating Casualties in a War to Overthrow Saddam	*Orbis*
2004	Glosny	Strangulation from the Sea? A PRC Submarine Blockade of Taiwan	*International Security*

2004	Kuperman	*The Limits of Humanitarian Intervention: Genocide in Rwanda*	Brookings
2005	Armstrong and Powell	A Stochastic Salvo Model Analysis of the Battle of the Coral Sea	*Military Operations Research*
2006	Lieber and Press	The End of MAD? The Nuclear Dimension of U.S. Primacy	*International Security*
2007	Raas and Long	Osirak Redux? Assessing Israeli Capabilities to Destroy Iranian Nuclear Facilities	*International Security*
2007	Bjoern	African Adventure?: Assessing the European Union's Military Intervention in Chad and the Central African Republic	MIT Security Studies
2008	Talmadge	Closing Time: Assessing the Iranian Threat to the Strait of Hormuz	*International Security*
2010	Gholz and Press	Protecting "The Prize": Oil and the US National Interest	*Security Studies*
2011	Bennett and Lind	The Collapse of North Korea: Military Missions and Requirements	*International Security*
2011	Shifrinson and Priebe	A Crude Threat: The Limits of an Iranian Missile Campaign Against Saudi Arabian Oil	*International Security*
2012	Bell	Can Britain Defend the Falklands?	*Defence Studies*
2014	Haggerty	Safe Havens in Syria: Missions and Requirements for an Air Campaign	MIT Masters Thesis
2014	Anderson	Peacekeepers Fighting a Counterinsurgency Campaign: A Net Assessment of the African Union Mission in Somalia	*Studies in Conflict & Terrorism*
2015	Armstrong and Sodergren	Refighting Pickett's Charge: Mathematical Modeling of the Civil War Battlefield	*Social Science Quarterly*
2016	Heginbotham et al.	*The US-China military scorecard*	RAND
2016	Glaser and Fetter	Should the United States Reject MAD? Damage Limitation and Nuclear Strategy toward China	*International Security*
2016	Biddle and Oelrich	Future Warfare in the Western Pacific: Chinese Antiaccess/Area Denial, U.S. AirSea Battle, and Command of the Commons in East Asia	*International Security*
2016	MacKay, Price, and Wood	Weighing the Fog of War: Illustrating the Power of Bayesian Methods for Historical Analysis through the Battle of the Dogger Bank	*Historical Methods*
2017	Talmadge	Would China Go Nuclear? Assessing the Risk of Chinese Nuclear Escalation in a Conventional War with the United States	*International Security*

2020	Fagan, Horwood, MacKay, Price, Richards, and Wood	Bootstrapping the Battle of Britain	*Journal of Military History*
2020	Wu	Living with Uncertainty: Modeling China's Nuclear Survivability	*International Security*

Appendix B: Selecting Input Distributions

Our proposed approach to handling uncertainty requires researchers to select a distribution for the inputs, decide on its parameters, and to model the correlation between different inputs. Guidance for parameter assignment can be found in Step 4, and we provide guidance for distribution selection and correlation here. Although implementing Monte Carlo once distributions are selected is a computational task, substantial qualitative research (e.g. analysis of historical campaigns, consultations with experts), should inform distribution selection. Many statistical distributions are available for researchers to use. The choice of distribution can affect the outcome, both by producing the incorrect variance for the outcome measure, but also potentially by introducing statistical bias, though most of the simulation literature focuses on variance.[1] As with constructing the model, no universal rules exist for selecting the appropriate distribution.[2] That said, there is some general guidance on which distributions are appropriate in which circumstances. Researchers can draw on their substantive knowledge and conduct research (e.g. examining historical campaigns or consulting experts) to decide on distributions, or they could conduct an additional form of sensitivity analysis, comparing the results from using different distributions.[3]

The distribution that encodes the least information about a variable is a *uniform* distribution. If researchers can specify upper and lower bounds on a variable but have no other information about which values within that range are more likely than any other, then a uniform distribution is the appropriate choice. For instance, in their sensitivity analysis, Press and Lieber examine a range of nuclear weapon reliability and accuracy, plugging in values from uniform distributions.[4]

Researchers conducting campaign analysis often have an estimate of a most likely value, in addition to the upper and lower bounds of model variables. When researchers have a sense of these three values only, the simulations literature suggests using a *triangular* distribution, with a maximum value at the most likely value suggested by the expert, and probability decreasing linearly (and potentially asymmetrically) to 0 at each of the extremes.[5]

[1]See Eunhye Song and Barry L Nelson, "Input Model Risk," in, *Advances in Modeling and Simulation* (Springer, 2017), pp. 63–80, pp. 68-69

[2]Alan R Washburn and Moshe Kress, *Combat Modeling*, vol. 139 (Springer, 2009), p. 9.

[3]Bahar Biller and Barry L Nelson, "Answers to the Top Ten Input Modeling Questions," in, *Proceedings of the Winter Simulation Conference*, vol. 1 (IEEE, 2002), pp. 35–40, doi:10.1109/WSC.2002.1172865; Song and Nelson, "Input Model Risk."

[4]Keir A Lieber and Daryl G Press, "The End of MAD? The Nuclear Dimension of US Primacy," *International Security*, Vol. 30, No. 4 (2006), pp. 7–44, doi:10.1162/isec.2006.30.4.7.

[5]Biller and Nelson, "Answers to the Top Ten Input Modeling Questions.". Several R packages and Python's numpy library provide functions for sampling from a triangular distribution.

The parameters of a normal distribution are more difficult to set qualitatively and normal distributions are symmetrical and unbounded, but is a defensible choice for many variables. The mean of a set of independent variables, regardless of their distributions, is asymptotically normally distributed, making it a valid choice for some types of variables, such as the total time for an operation, given independent components that make up the total time. More concretely, firing errors are usually normally distributed, making normal distributions valuable for modeling weapon accuracy.

Other distributions have attractive statistical properties but have parameters that are difficult to extract in the research process. Log-normal or beta distributions are useful distributions for continuous outcomes that are positive and continuous (log-normal) or between 0 and 1 (beta). Setting the parameters for these will often involve data that may not be available. The number of independent events or arrivals in a time period are distributed as a Poisson distribution, but the "rate" parameter can be difficult to estimate qualitatively.[6] Finally, some specific problems have distributions that have been characterized in the operations research or military modeling literatures. Research in search theory, for instance, shows that the probability of finding an object, $P(S)$, if searching effort is randomly allocated within the search area, is an exponential distribution $P(S) = 1 - e^{(-\text{effort}/\text{area})}$.[7]

Researchers must also decide how to model the dependence between different inputs. A known pitfall in the combat modeling literature, which relies heavily on statistical distributions, is the assumption that all inputs are independently distributed.[8] A researcher might believe that if one variable takes a high value, a second variable is also likely to take a high value. For instance (as we discuss further in the replication section), the probability that the United States would detect one type of Chinese nuclear facility could be correlated to some degree with the probability of detecting another type of nuclear facility. The simplest way to impose correlation is to use perfect correlation for all values: the detection probabilities for all targets might be identical within a single run. Often, though, researchers would not like to assume perfect correlation. If researchers are using normal distributions, they can set a level of covariance between them and draw from a multivariate normal.[9]

[6] This is nothing to say of events often being correlated, which makes the Poisson distribution inappropriate.
[7] Bernard Osgood Koopman, "Search and Screening," *OEG Rep.*
[8] Washburn and Kress, *Combat Modeling*.
[9] Textbooks on Bayesian hierarchical modeling also have useful guidance on creating joint distributions.