The Case for Campaign Analysis

A Method for Studying Military Operations

Rachel Tecott and Andrew Halterman

Could NATO stop 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. These questions are also difficult to tackle given the complexities and uncertainties of combat, and most efforts occur

Rachel Tecott is a Ph.D. candidate in political science at the Massachusetts Institute of Technology and a member of its Security Studies Program. Andrew Halterman is a Ph.D. candidate in political science at the Massachusetts Institute of Technology and a member of its Security Studies Program and Political Methodology Lab.

1. John J. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe," *International Security*, Vol. 7, No. 1 (Summer 1982), pp. 3–39, doi.org/10.2307/2538686; Joshua M. Epstein, *Measuring Military Power: The Soviet Air Threat to Europe*, Vol. 199 (Princeton, N.J.: Princeton University Press, 1984); Barry R. Posen, "Measuring the European Conventional Balance: Coping with Complexity in Threat Assessment," *International Security*, Vol. 9, No. 3 (Winter 1984/85), pp. 47–88, doi.org/10.2307/2538587; Joshua M. Epstein, "Dynamic Analysis and the Conventional Balance in Europe," *International Security*, Vol. 12, No. 4 (Spring 1988), pp. 154–165, doi.org/10.2307/2538999; and Barry R. Posen, "Is NATO Decisively Outnumbered?" *International Security*, Vol. 12, No. 4 (Spring 1988), pp. 186–202, doi.org/10.2307/2539002.
2. Keir A. Lieber and Daryl G. Press, "The End of MAD? The Nuclear Dimension of U.S. Primacy," *International Security*, Vol. 30, No. 4 (Spring 2006), pp. 7–44, doi.org/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 (Summer 2011), pp. 167–201, doi.org/10.1162/ISEC_a_00048; Caitlin Talmadge, "Closing Time: Assessing the Iranian Threat to the Strait of Hormuz," *International Security*, Vol. 33, No. 1 (Summer 2008), pp. 82–117, doi.org/10.1162/isec.2008.33.1.82; and Eugene Gholz and Daryl G. Press, "Protecting 'The Prize': Oil and the U.S. National Interest," *Security Studies*, Vol. 19, No. 3 (2010), pp. 453–485, doi.org/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 (Fall 1996), pp. 139–179, doi.org/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 (Spring 2017), pp. 50–92, doi.org/10.1162/ISEC_a_00274.

within governments, think tanks, and government-funded research centers. Scholars, however, have a 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: (1) formulating a question, (2) specifying a scenario, (3) constructing a model that represents the military operation, (4) setting values for those variables using qualitative research and technical military information, (5) running the model with sensitivity analysis, and (6) interpreting the output of the model and presenting the conclusions of the analysis.[6]

Security studies is in the midst of a methodological renaissance. 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 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 also have developed sophisticated methods for measuring core security studies concepts, such as territorial control and perception of threats and signals.[9]

Campaign analysis has not yet received such treatment. Although scholars have used the method for decades, campaign analysis remains underspecified in the academy, and techniques for conducting it are an oral tradition among a small number of scholars. This article 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

---

6. We use the term "campaign analysis" because it is the term most scholars currently use to describe the method we formalize here. As we discuss below, researchers can use the method to examine engagements from the campaign down to the tactical levels of conflict.
7. Christopher Darnton, "Archives and Inference: Documentary Evidence in Case Study Research and the Debate over U.S. Entry into World War II," *International Security*, Vol. 42, No. 3 (Winter 2017/18), pp. 84–126, doi.org/10.1162/ISEC_a_00306.
8. Erik Lin-Greenberg, "Wargame of Drones: Remotely Piloted Aircraft and Crisis Escalation," *SSRN*, working paper, Massachusetts Institute of Technology, 2020, doi.org/10.2139/ssrn.3288988; Reid B.C. Pauly, "Would U.S. Leaders Push the Button? Wargames and the Sources of Nuclear Restraint," *International Security*, Vol. 43, No. 2 (Fall 2018), pp. 151–192, doi.org/10.1162/isec_a_00333; and Elizabeth M. Bartels, "Building Better Games for National Security Policy Analysis," Ph.D. thesis, Pardee RAND Graduate School, 2020, doi.org/10.7249/RGSD437.
9. Therese Anders et al., "Measuring Territorial Control in Civil Wars Using Hidden Markov Models: A Data Informatics-Based Approach," paper presented at the 31st Conference on Neural Information Processing Systems Workshop on Machine Learning for the Developing World, Long Beach, California, December 8, 2017, https://arxiv.org/pdf/1711.06786.pdf; Marika Landau-Wells, "Dealing with Danger: Threat Perception and Policy Preferences," Ph.D. thesis, Massachusetts Institute of Technology, 2018; and Azusa Katagiri and Eric Min, "The Credibility of Public and Private Signals: A Document-Based Approach," *American Political Science Review*, Vol. 113, No. 1 (February 2019), pp. 156–172, doi.org/10.1017/S0003055418000643.

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 available. 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 contribute to scholarly inquiry by revealing theoretical puzzles, suggesting new theories, and producing alternative measures for key variables in theoretical debates.

Critics have raised several objections to campaign analysis as a method for studying military operations. Scholars have argued that it leaves out political-economic-social variables and focuses only on one or a handful out of the many possible scenarios that could shape conflict.[10] 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 Barry Posen observes, specialized attention to individual scenarios is a necessary step toward 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."[11]

Campaign analysis also has 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.[12] More broadly, war is too complex and uncertain to model.[13] Military operations are indeed difficult to model well, and many critical variables cannot be estimated with precision. Academic researchers, however, not only are familiar with military operations and the political conditions that shape them, but they also are equipped with

10. Eliot A. Cohen, "Toward Better Net Assessment: Rethinking the European Conventional Balance," *International Security*, Vol. 13, No. 1 (Summer 1988), pp. 50–89, doi.org/10.2307/2538896.
11. Posen in John J. Mearsheimer, Barry R. Posen, and Eliot A. Cohen, "Correspondence: Reassessing Net Assessment," *International Security*, Vol. 13, No. 4 (Spring 1989), pp. 128–179, at p. 146, doi.org/10.2307/2538782.
12. Cohen, "Toward Better Net Assessment," especially pp. 59–60, 85.
13. Carl von Clausewitz, *On War*, trans. and ed. Michael Howard and Peter Paret (Princeton, N.J.: Princeton University Press, 1976); and Barry D. Watts, "Ignoring Reality: Problems of Theory and Evidence in Security Studies," *Security Studies*, Vol. 7, No. 2 (Winter 1997/98), pp. 115–171, doi.org/10.1080/09636419708429344.

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 article is explicitly designed to facilitate valid inference in the face of uncertainty. Although campaign analysis does not enable researchers to perfectly explain or predict every aspect of military operations, if well done, it does equip them to answer questions of great value to policy and academic research.

The purpose of this article is to encourage researchers to use campaign analysis to study the military operations at the very center of international relations theory and practice. In service of this objective, it provides a methodological toolkit for researchers and illustrates the value of the method for theory and policy.

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. Fourth, we replicate and extend two published campaign analyses—Wu Riqiang's analysis of Chinese nuclear survivability and Barry Posen's analysis of NATO's prospects against the Warsaw Pact—highlighting the benefits of the two recommendations and the value of the method for academic theory. We conclude with a summary of our main arguments and discussion of future campaign analysis research.

Campaign Analysis—A Distinct Method

Throughout this article, we refer to campaign analysis as a "method," a term that does not have a strict definition in social science research and is often used interchangeably with the terms "research design," "technique," or "tool."[14] Nevertheless, a method in social science research as commonly understood is a relatively structured approach to producing findings from data.[15] The methods that researchers employ depend on their questions,[16] and on the problems they need to solve to answer those questions. Social science research methods tend to have several common features: a set of questions or tasks for which they are

---

14. See, for example, Darnton, "Archives and Inference."

15. As Giovanni Sartori puts it, methodology "is a concern with the logical structure and procedure of scientific enquiry." Sartori, "Concept Misformation in Comparative Politics," *American Political Science Review*, Vol. 64, No. 4 (December 1970), pp. 1033–1053, at p. 1033, doi.org/10.2307/1958356.

16. Margaret E. Roberts, "What Is Political Methodology?" *PS: Political Science & Politics*, Vol. 51, No. 3 (July 2018), pp. 597–601, doi.org/10.1017/S1049096518000537.

useful and appropriate; an approach to solving a specific kind of inferential problem; and a set of shared standards, such that a reader can evaluate whether the method has been appropriately employed for valid inference.

DISTINGUISHING CAMPAIGN ANALYSIS FROM RELATED METHODS

Campaign analysis is one of many methods that academic, government, and military researchers have developed to understand conflict. Some techniques, including wargames, tabletop exercises, and field exercises, stretch back thousands of years.[17] Others took root during World War II and over the course of the Cold War. These methods include the (sometimes overlapping) fields of operations research, systems analysis, modeling and simulations, game theory, and net assessment.

During World War II, the Allied governments began using operations research, which draws on applied mathematics to improve tactical and operational force employment decisions. In particular, operations research informed Allied antisubmarine warfare, bombing techniques, and submarine tactics.[18] Since then, the field of operations research has developed new mathematical tools for modeling military operations and, outside the military context, for deriving the optimal allocation of resources in a wide range of scenarios. Scholars conducting operations research today often focus on abstract approaches to solving general classes of problems, rather than on deep examination of specific military scenarios. Military operations research tends to address optimization problems such as how to optimally allocate search efforts when looking for a target,[19] or how to optimally allocate warheads to targets,[20] whereas the broader operations research literature (in computer science and business schools) has studied a wide range of constrained optimization problems.[21]

---

17. Paul K. Davis, "Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation," *Proceedings of the IEEE*, Vol. 83, No. 8 (August 1995), pp. 1138–1155, doi.org/10.1109/5.400454; and Roger D. Smith, "Essential Techniques for Military Modeling and Simulation," in *1998 Winter Simulation Conference. Proceedings (Cat. No. 98CH36274)*, Washington, D.C., 1998, Vol. 1, pp. 805–812, doi.org/10.1109/WSC.1998.745067.
18. Philip McCord Morse and George E Kimball, *Methods of Operations Research*, 1951, ed. (Mineola, N.Y.: Dover, 2003).
19. Henry R. Richardson and Lawrence D. Stone, "Operations Analysis during the Underwater Search for Scorpion," *Naval Research Logistics Quarterly*, Vol. 18, No. 2 (June 1971), pp. 141–157, doi.org/10.1002/nav.3800180202; Lawrence D. Stone, *Theory of Optimal Search* (Amsterdam: Elsevier, 1976); and Lawrence D. Stone et al., *Optimal Search for Moving Targets* (Cham, Switzerland: Springer, 2016).
20. Ravindra K. Ahuja et al., "Exact and Heuristic Algorithms for the Weapon-Target Assignment Problem," *Operations Research*, Vol. 55, No. 6 (November–December 2007), pp. 1136–1146, doi.org/10.1287/opre.1070.0440.
21. For more recent general military operations research techniques, see Mike Cornforth and Wayne P. Hughes Jr., *Military Modeling for Decision Making*, 3d ed. (Arlington, Va.: Military Operations Research Society, 1997); and Alan R. Washburn and Moshe Kress, *Combat Modeling*, Vol. 139 (New York: Springer, 2009). For applied civilian operations research, see the efficiency of the single checkout line at Trader Joe's.

Closely related to operations research, "systems analysis" and "modeling and simulations" methods use applied models to assist military commanders with a wide range of 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 equipment, expert opinion, and historical experience, and often integrate many submodels into larger, multi-resolution models. The most sophisticated models, such as the STORM (Synthetic Theater Operations Research Model) or RSAS (RAND Strategy Assessment System) models used in the U.S. Defense Department, can help commanders plan for everything from the tactical outcomes of armored battles on a specific piece of real-world terrain to the projected use of spare parts during air operations.[22]

Campaign analysis evolved from and remains closely related to operations research, systems analysis, and modeling and simulations methods. It is distinct from these in their ideal types, however, by (1) the kinds of questions it answers, (2) the complexity of its models, (3) the data used to assign model parameters, and (4) its approach to managing uncertainty.

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 prizes 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.

Campaign analysis models tend to be much simpler than modeling and simulations models, which are usually highly complex, incorporating hundreds or thousands of variables to create a tool that can be applied to different problem sets of different potential users. Because researchers employing campaign analysis ask only one or several questions under carefully specified scenarios, they can use simpler models than the models required to answer many potential questions at different layers of resolution. The simplicity of campaign analysis models is not a weakness: complex models (and their implementation in code) can be opaque, conceal biases, and behave in unexpected ways. The complexity of some modeling and simulations provides great benefit in studying the detailed components of military operations, but their complexity, growing out of their specific questions, should not dissuade scholars from constructing simple models, which have their own advantages.

The approaches also differ in the extent to which they use values or data.

---

22. For a history of Defense Department modeling and simulation efforts see Davis, "Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation." See, for example, TACWAR. Robert J. Atwell and D. Graham McBryde, "Theater-Level Ground Combat Analyses and the TACWAR Submodels" (Alexandria, Va.: Institute for Defense Analyses, 1991).

Many operations research articles focus on developing abstract models, rather than on plugging values into model parameters to produce an answer.[23] Military models and simulations, like campaign analysis, assign estimated real-world values to model parameters to produce predictions for specific scenarios.[24] The large number of variables in many military models and simulations demands a great deal of empirical research to set all of the model values, whereas campaign analysis models often contain fewer variables and are thus more manageable for a single scholar or small team to research. Also, campaign analysis conducted by academics usually relies on open-source data to assign parameter values, whereas operations research and campaign analysis conducted for the government can draw on classified data.

Finally, campaign analysis is unique in its approach to managing uncertainty. It is frequently used to answer questions about hypothetical military operations where data are scarce and sometimes classified. Campaign analysis manages uncertainty through the careful construction and defense of a simple model, assignment of parameter values, and sensitivity analysis.

Campaign analysis also is related to, yet distinct from, game theory and formal modeling. Game theory, the study of rational decisionmaking in strategic interactions, was also used during World War II,[25] and it fundamentally shaped defense thinking and scholarship during the Cold War.[26] Game theorists identify the actors, structure the game, state assumed actor preferences, and then mathematically derive the rational choices of the actors in abstract circumstances. Game theory is a subset of formal modeling, a method that has been used to understand a wide range of problems in security studies and political science. Although 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 pioneered by Andrew Marshall to help the U.S. government plan for long-term competition with the Soviet Union.

---

23. For example, we start by assuming the target is stationary at the point $x = (x_1, x_2)$.
24. For example, we start with a Russian Akula-class submarine stationary at the coordinates $(35.948, -5.574)$.
25. Morse and Kimball, *Methods of Operations Research*; and John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton, N.J.: Princeton University Press, 1953).
26. See, perhaps most prominently, research on nuclear deterrence, including the seminal Thomas C. Schelling, *Arms and Influence* (New Haven, Conn.: Yale University Press, 1966).

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.[27] Net assessment is thus an expansive research agenda.

In contrast, researchers use campaign analysis to tackle much more discrete projects. Rather 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 pages of *International Security* in the 1980s. Substantive debate between NATO "pessimists" and NATO "optimists" evolved into discussion of the feasibility and value of academic campaign analysis. Eliot Cohen, with a background in net assessment, criticized Posen and John 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 on 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.[28]

CAMPAIGN ANALYSIS—A BIG TENT

Although campaign analysis is distinct from other methods of military science, the method remains a "big tent," suitable for answering a wide range of questions. Scholars can use campaign analysis to study various levels and all kinds of warfare, and to answer questions about both hypothetical conflict and historical conflict.

Campaign analysis can be applied to questions below the strategic level of war. The terms "operations" and "campaigns" are sometimes used interchangeably, with operations generally understood to comprise campaigns. The

---

27. Thomas G. Mahnken, ed., *Net Assessment and Military Strategy: Retrospective and Prospective Essays* (Amherst, N.Y.: Cambria, 2020).
28. Other researchers have also used the term "net assessment" to refer to methods that fit our definition of campaign analysis.

U.S. Defense Department defines a "campaign" as "a series of related operations aimed at achieving strategic and operational objectives within a given time and space."[29] It defines an "operation" as "1. A sequence of tactical actions with a common purpose or unifying theme. (Joint Publication 1) 2. A military action or the carrying out of a strategic, operational, tactical, service, training, or administrative military mission. (Joint Publication 3-0)."[30] The Defense Department further defines the "operational level of warfare" as "the level of warfare at which campaigns and major operations are planned, conducted, and sustained to achieve strategic objectives within theaters or other operational areas. See also strategic level of warfare; tactical level of warfare. (Joint Publication 3-0)."[31] Although we use the term "campaign analysis," the method does not exclude tactical or operational questions.[32]

Campaign analysis has most often been used to answer questions about the operational level of warfare. Operational outcomes are often of great interest to political scientists: whether or not a state's military could execute a fait accompli, defend against an armored breakthrough, or close a body of water to shipping are operational-level questions of tremendous importance because of the roles these operations have played in historical conflicts and international relations theories of war, deterrence, and coercion. In contrast, tactical engagements on their own rarely have the same political importance of many operations. Political scientists are unlikely to be interested, for instance, in determining the optimal processes for launching aircraft from a carrier or the best employment of helicopters for raiding missions.[33] On the opposite end of the spectrum, questions such as "How can the United States maintain its military advantage over China over the next thirty years?" are better addressed through net assessment and other approaches.

Campaign analysis can be used to study any kind of warfare, including conventional, nuclear, and unconventional warfare. The method of campaign analysis took root in academic security studies in the 1980s in the context of efforts to assess the conventional balance of forces in Western Europe.[34] Since the

---

29. Joint Publication 5-0, "DOD Dictionary of Military and Associated Terms," https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf, p. 159.
30. Ibid.
31. Ibid., p. 161.
32. Hughes presents an alternative taxonomy. Wayne P. Hughes, "Overview," in Hughes, ed., *Military Modeling for Decision Making* (Arlington, Va.: Military Operations Research, 1989).
33. There are exceptions, however: some "tactical" engagements, such as individual missile strikes on targets, might be highly consequential and are relatively simple to model. See, for example, Lieber and Press, "The End of MAD?"; and Shifrinson and Priebe, "A Crude Threat."
34. 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"; and Barry R. Posen, *Inadvertent Escalation: Conventional War and Nuclear Risks* (Ithaca, N.Y.: Cornell University Press, 1991). Researchers in academic security studies had been using similar techniques since the 1970s to study nuclear conflict,

fall of the Berlin Wall, researchers have employed campaign analysis to answer questions about conventional war on the Korean Peninsula,[35] humanitarian operations,[36] nuclear operations,[37] air power and missile strikes,[38] counterinsurgency,[39] and a wide variety of military operations.[40] Although some types of questions, such as those regarding counterinsurgency, may involve more uncertainty and difficulty in modeling, nothing in the definition of campaign analysis excludes counterinsurgency.

Campaign analysis can be used to address questions about hypothetical or historical military operations. Historical campaign analyses often involve the study of counterfactuals, such as how changing antisubmarine warfare tactics could have saved more Allied shipping or how adding a U.S. carrier might have changed the outcome of the Battle of the Coral Sea.[41] Historians, mathematicians, and political scientists have employed campaign analysis to examine the outcomes of historical engagements including Pickett's Charge and the Battle of the Dogger Bank, and campaigns including the Battle of Britain and Operation Barbarossa.[42]

---

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 (June 1973), pp. 207–242, doi.org/10.1177/002200277301700203; and John D. Steinbruner and Thomas M. Garwin, "Strategic Vulnerability: The Balance between Prudence and Paranoia," *International Security*, Vol. 1, No. 1 (Summer 1976), pp. 138–181, doi.org/10.2307/2538581.

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

36. 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.org/10.1080/714005314; and Alan J. Kuperman, *The Limits of Humanitarian Intervention: Genocide in Rwanda* (Washington, D.C.: Brookings Institution Press, 2004).

37. Lieber and Press, "The End of MAD?"

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

39. Bjoern H. Seibert, "African Adventure?: Assessing the European Union's Military Intervention in Chad and the Central African Republic" (Cambridge, Mass.: Security Studies Program, Massachusetts Institute of Technology, 2007).

40. Talmadge, "Closing Time"; and Mark S. Bell, "Can Britain Defend the Falklands?" *Defence Studies*, Vol. 12, No. 2 (June 2012), pp. 283–301, doi.org/10.1080/14702436.2012.699726.

41. Brian McCue, *U-Boats in the Bay of Biscay: An Essay in Operations Analysis* (Washington, D.C.: National Defense University Press, 1990); and Michael J. Armstrong and Michael B. Powell, "A Stochastic Salvo Model Analysis of the Battle of the Coral Sea," *Military Operations Research*, Vol. 10, No. 4 (August 2005), pp. 27–37, doi 10.5711/morj.10.4.27. For another example of historical campaign analysis, see Biddle, "Victory Misunderstood."

42. 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 (December 2015), pp. 1153–1168, doi.org/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, doi.org/10.1080/01615440.2015.1072071; Brennen Fagan et al., "Bootstrapping the Battle of Britain," *Journal of Military History*, Vol. 84, No. 1 (January

Because the outcomes of historical campaigns are observable, the values of key variables can often be recovered, and processes linking variables to outcomes can be traced, researchers interested in explaining why historical military operations unfolded the way that they did have a wide range of methodological approaches available to them in addition to campaign analysis. For example, Stephen Biddle uses process tracing—not campaign analysis—to examine the causes of U.S. success in the invasion of Iraq in 2003.[43] The availability of historical data to inform their study of the engagement means that researchers do not need to rely on a model.[44] Otherwise put, campaign analysis is a useful method for examining historical operations, but it is just one of many methods that researchers can use to examine past military operations.

Hypothetical military campaigns, however, have no actual outcome for researchers to observe, and many of the variables that compose the models are likewise unobservable. A researcher seeking to determine whether NATO could have held off a Warsaw Pact armored invasion cannot look to history for the results of any actual engagement between NATO and Soviet forces, nor can that researcher observe the values of key variables, such as the time it might have taken NATO to mobilize. Consequently, hypothetical military operations cannot be studied with familiar methods such as interviews with combatants or analysis of military after-action reports, and methods such as campaign analysis are especially helpful.

Campaign analysis is thus a method in social science, and one distinct from related "military science" methods such as net assessment or operations research, because researchers use campaign analysis to address different kinds of questions and employ a distinct approach to the management of complexity and uncertainty to answer those questions. Yet unlike many methods in security studies, methodological guidance for campaign analysis is sparse. Some researchers intersperse methodological tidbits within their substantive projects, but none have offered a comprehensive discussion. For instance, Mearsheimer gives high-level guidance on conducting a "theater balance" beginning with the actors and their preferences, and presents some advice on how to assess the conventional balance in Europe.[45] Charles Kupchan argues for careful approaches to quantification, sensitivity analysis, and think-

2020), pp. 151–186; and Ryan T. Baker, "Logistics and Military Power: Tooth, Tail, and Territory in Conventional Military Conflict," Ph.D. thesis, George Washington University, 2020.

43. 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), p. 8, doi.org/10.1080/01402390701210749.

44. Models can be useful for historical campaigns as well. McCue, *U-Boats in the Bay of Biscay*.

45. John J. Mearsheimer, "Numbers, Strategy, and the European Balance," *International Security*, Vol. 12, No. 4 (Spring 1988), pp. 174–185, doi.org/10.2307/2539001.

ing about clear "yardsticks of sufficiency" in analyzing military operations.[46] Joshua Epstein briefly discusses the use of a model and management of uncertainty to answer questions about hypothetical military operations.[47]

No previous work, however, has defined the method of campaign analysis, standardized it, or offered comprehensive guidance for how to conduct and evaluate it. The next section provides this methodological guidance.

Six Steps of Campaign Analysis

In this section, we standardize the six core steps of campaign analysis and provide guidance for how to conduct them for valid inference.[48]

STEP 1: FORMULATE A QUESTION

Campaign analysis, like all methods in social science, is intended to answer motivating research questions. It 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 destroy Saudi oil refineries or close the Strait of Hormuz?[49] Could NATO conventional forces defend Western Europe against a Soviet attack?[50] These questions are often framed as "sufficiency" questions: Is a campaign achievable, impossible, or is the outcome indeterminate? Charles Glaser and Chaim Kaufmann 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?"[51] Researchers also can use campaign analysis to examine the effects of a specific variable on the outcome of hypothetical military operations. For instance, Wu asks how the level of dispersion of Chinese nuclear forces changes their survivability.[52]

---

46. Charles A. Kupchan, "Setting Conventional Force Requirements: Roughly Right or Precisely Wrong?" *World Politics*, Vol. 41, No. 4 (July 1989), pp. 536–578, doi.org/10.2307/2010529. For "yardsticks of sufficiency," see Alain C. Enthoven and K. Wayne Smith, *How Much Is Enough?: Shaping the Defense Program, 1961–1969* (Santa Monica, Calif.: RAND Corporation, 1971).
47. Epstein, *Measuring Military Power*.
48. Although we call them steps, the process of conducting campaign analysis is often iterative.
49. Shifrinson and Priebe, "A Crude Threat"; and Talmadge, "Closing Time."
50. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe"; and Posen, "Is NATO Decisively Outnumbered?"
51. Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?" p. 75.
52. Wu Riqiang, "Living with Uncertainty: Modeling China's Nuclear Survivability," *International Security*, Vol. 44, No. 4 (Spring 2020), pp. 84–118, doi.org/10.1162/isec_a_00376.

As in the rest of social science, there are no methodological rules to guide question selection.[53] Broadly speaking, researchers tend to choose questions based on interest and tractability. 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.[54] 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 "What is 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 accomplis 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 influences the questions that campaign analysis researchers pursue. For instance, we might be interested in how the B-2 stealth bomber affected the U.S. 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.

Researchers conducting campaign analysis then move from motivating questions to specific outcomes of interest. A general question, such as whether

---

53. Stephen Van Evera, *Guide to Methods for Students of Political Science* (Ithaca, N.Y.: Cornell University Press, 1997).
54. See, as one exception, Wu, "Living with Uncertainty."

the United States military could have mitigated the genocide in Rwanda, becomes a specific question about how quickly different configurations of U.S. forces could have arrived, and an outcome of how many days it would have taken for a given force to arrive.[55] 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: Alan Kuperman had to first defend his argument that the time it would have taken U.S. 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" (how many German planes were shot down by merchant ships' anti-aircraft guns) can obscure better measures of effectiveness (how many Allied ships survived crossing the Atlantic).[56] Joshua Shifrinson and Miranda Priebe are interested in whether an Iranian missile strike could disable Saudi oil production.[57] After determining that Saudi pipelines could be easily repaired, they chose to estimate Iran's capacity to destroy the Abqaiq oil stabilization facility, a much more important bottleneck.

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.[58] Sufficiency outcomes often provide the answers that scholars want, and they are often much easier to model in the context of great uncertainty than precise point estimate outcomes because more variables can be safely omitted. 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: SPECIFY A 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

---

55. Kuperman, *The Limits of Humanitarian Intervention*, p. 66.
56. Morse and Kimball, *Methods of Operations Research*, pp. 52–53.
57. Shifrinson and Priebe, "A Crude Threat."
58. Talmadge, "Closing Time."

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 vary them to see how military operations would 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, or 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 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.[59] Whereas researchers have not frequently 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 is one that 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 un-

---

59. Wu, "Living with Uncertainty," p. 86.

likely, selecting a "hard" scenario can make their conclusion more robust. For example, Mark 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.[60] Researchers also have 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.[61] 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, states 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.[62]

STEP 3: CONSTRUCT A 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."[63] 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. Sec-

---

60. Bell, "Can Britain Defend the Falklands?"
61. Lieber and Press, "The End of MAD?"
62. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe."
63. 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.

ond, 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. As Caitlin 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."[64]

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.[65] 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.[66] 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. Michael O'Hanlon uses this validation technique by studying a North Korean invasion of South Korea, using both a theater-level model and a zoomed-in model of a specific battle, reaching similarly pessimistic conclusions about North Korea's prospects.[67] 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 inputs change.[68] 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.[69] 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 re-

---

64. Talmadge, "Closing Time," p. 84.
65. James S. Hodges et al., *Is It You or Your Model Talking? A Framework for Model Validation* (Santa Monica, Calif.: RAND Corporation, 1992).
66. Robert G. Sargent, "Verification and Validation of Simulation Models," *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, Maryland, 2010, pp. 166–183, doi.org/10.1109/WSC .2010.5679166.
67. O'Hanlon, "Stopping a North Korean Invasion."
68. Sargent, "Verification and Validation of Simulation Models."
69. Ibid.

searcher 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, Paul 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."[70]

In general, the model should be the simplest model that provides "useful" results for the question being asked.[71] Simple models have four advantages. First, more complicated models are more resource intensive.[72] It is often easier to assign plausible value ranges to high-level aggregate variables (e.g., average convoy speed) than to subvariables (e.g., speed of a specific type of tank across 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 bounds 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 (Tactical Warfare) military modeling program found more than seventy programming errors, including failures to reset counters, division by zero errors, and errors in how aircraft are allocated to close air support or attacks on enemy air bases.[73] 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.[74]

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

70. Davis, "Distributed Interactive Simulation in the Evolution of DoD Warfare Modeling and Simulation," p. 1139.
71. 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 quotation in their first semester of graduate school about the idea that "all models are wrong, but some are useful." George E.P. Box, "Science and Statistics," *Journal of the American Statistical Association*, Vol. 71, No. 356 (1976), pp. 791–799, doi.org/10.2307/2286841.
72. Stewart Robinson, "Tutorial: Choosing What to Model–Conceptual Modeling for Simulation," *Proceedings of the 2012 Winter Simulation Conference*, Berlin 2012, pp. 1–12, doi.org/10.1109/WSC.2012.6465308.
73. John C. Ingram, "A Detailed Review of the TACWAR Model" (Adelphi, Md.: Harry Diamond Laboratories, 1980).
74. Robinson, "Tutorial Choosing," p. 1916.

communicate the model's results clearly to readers.[75] 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 exclude from and which to include in 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.

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 state (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" (SSPK) against a target, meaning that the model should not then also include the (now redundant) components of SSPK: accuracy, yield, and target hardness.[76]

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 they would favor NATO,[77] only strengthening his argument.[78]

75. See O'Hanlon, "Stopping a North Korean Invasion," p. 154.
76. Davis and Schilling, "All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask."
77. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe," p. 4 n. 5.
78. This reason depends on a broad agreement among readers in the direction of the effect. Cohen challenged Mearsheimer's assessment of airpower as favorable to NATO.

There are unsound reasons to exclude model variables as well. Alan Washburn and Moshe 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.[79] If a researcher discovers that 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 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 argument.

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.[80] 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.[81] For instance, the physics of ballistic missiles and nuclear destruction are well summarized in the open source literature.[82] Keir Lieber and Daryl Press are interested in the

---

79. Washburn and Kress, *Combat Modeling*.

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

81. The military operations research literature offers a starting point for model construction. See the journals *Operations Research* and *Military Operations Research* and the proceedings of the Winter Simulation Conference. For 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, Calif.: RAND Corporation, 1996), p. 13.

82. Samuel Glasstone, *The Effects of Nuclear Weapons* (Germantown, Md.: U.S. Atomic Energy Commission, 1962); and Davis and Schilling, "All You Ever Wanted to Know About MIRV and ICBM Calculations but Were Not Cleared to Ask."

effects that increasing warhead accuracy and missile reliability have on U.S. 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.[83]

Researchers often construct models of hypothetical military operations by examining historical operations. Whitney Raas and Austin 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 Iraq's Osirak nuclear reactor in 1981.[84] 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 prevail.[85]

Researchers constructing models of historical military operations have additional options. Brian 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.[86] McCue extends Philip Morse and George 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 German Adm. Karl Dönitz's wartime diaries, to develop his more sophisticated model. Specifically, he models the repair capacity of shipyards 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. For example, Talmadge studies a hypothetical U.S. counter-mine operation in the Persian Gulf.[87] Talmadge is interested in whether Iranian naval forces could close the Strait of Hormuz with mines, so she develops a simple

---

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

84. Raas and Long, "Osirak Redux?"

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

86. McCue, *U-Boats in the Bay of Biscay*.

87. Talmadge, "Closing Time."

model of naval minesweeping that posits that each minesweeping vessel can clear a certain number of mines per day of operation.[88] Talmadge could have chosen to draw on the military planning literature on mine effects and mine clearing to develop a very high-resolution model of where contact and influence mines might be laid, the clearing patterns of minesweeping vessels, and the probability of activating either contact or influence mines on each pass.[89] Such model complexity would have increased the difficulty of the research unnecessarily, introduced more opportunities for error in setting parameters, and made her analysis less communicable to her intended audience.

STEP 4: ASSIGN VALUES

Once the model is specified, researchers must assign values to each parameter to answer the question that motivates the study. Uncertainty in parameter values is a critical challenge for campaign analysis, and value assignment 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 such as 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 hard 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."[90]

---

88. Ibid.

89. Alan R. Washburn, "Mine Warfare Models" (Monterey, Calif.: Naval Postgraduate School, 2007).

90. Epstein, *Measuring Military Power*, pp. 199, xxviii.

Similarly, Bell defends his conclusion that Britain could defend the Falklands by selecting values favorable to Argentina.[91] For instance, he posits that British reinforcements would begin arriving within thirty-six hours of an Argentine attack, a conservative assessment against the British given that expert opinion assessed that they would arrive within twenty-four hours. If it holds even in a worst-case scenario, the conclusion will stand for any plausible input. This approach is therefore very powerful for researchers making claims about (in)sufficiency outcomes.

Practically, researchers can consult a number of sources 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 such as Jane's, the International Institute for Strategic Studies *Military Balance*, or think tank reports. Data on the order of battle or locations of forces are increasingly available using open source intelligence techniques.[92] 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.[93] For instance, in studying a potential North Korean invasion of South Korea, O'Hanlon draws on a U.S. Army study of historical rates of advance for armored units.[94] Data on historical operations are sometimes available in academic military histories, military after action reports, or even congressional testimony, which Kuperman uses to estimate the speed of U.S. strategic airlift.[95] In the best case, a previous analysis will be available, say from the RAND Corporation, 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 conser-

---

91. Bell, "Can Britain Defend the Falklands?"

92. For example, several researchers have used Google Earth satellite imagery to inform campaign analysis. Shifrinson and Priebe, "A Crude Threat"; and Decker Eveleth, "Mapping the People's Liberation Army Rocket Force," https://www.aboyandhis.blog/post/mapping-the-people-s-liberation-army-rocket-force.

93. Jacob A. Stockfisch offers useful questions for analysts examining input values. Stockfisch, "Models, Data, and War: A Critique of the Study of Conventional Forces" (Santa Monica, Calif.: RAND Corporation, 1975), p. vii.

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

95. Kuperman, *The Limits of Humanitarian Intervention*.

vative 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 for the final outcome to be accurate. The "most plausible" approach is often impractical, however, because available data are 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 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 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 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. They can re-run analysis multiple times with multiple model parameter values (basic sensitivity analysis conducted in existing research), or they can 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.[96] In practice and given space constraints, researchers often do not conduct sensitivity analysis for variables that

---

96. For example, consider our statement "Our results are robust to France having two aircraft carriers instead of one."

they believe have little effect on the final outcome, but this can leave the research open to criticism.

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.[97] Other researchers identify several key parameters and run the model multiple times using different sets of parameters.[98]

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. Whereas 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 that produce a given output does not obviate the need to improve confidence levels and to include error terms with all assessments."[99]

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 percent," 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 percent") or in qualitative terms ("a nuclear weapon will probably not survive").[100]

---

97. Shifrinson and Priebe, "A Crude Threat," p. 191.
98. See, for example, Posen, *Inadvertent Escalation*.
99. Kupchan, "Setting Conventional Force Requirements," p. 572.
100. Scholars and practitioners of intelligence 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, https://web.archive.org/web/20201024110519/https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol2no4/html; 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 (Fall 2017), pp. 803–826, doi.org/10.1017/S0020818317000352; and 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 (June 2018), pp. 410–422, doi.org/10.1093/isq/sqx078.

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."[101] 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."[102] 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."[103] From an ethical perspective, researchers should be attentive to how the findings might be interpreted or used by decisionmakers to justify different policies.

Two Recommendations for Researchers

We propose two recommendations for improving the method outlined above. Both recommendations would help researchers make better use of the qualitative research they have already conducted and better manage uncertainty.

RECOMMENDATION 1: PUBLISH A TRANSPARENT MODEL
One of the important intellectual contributions of a campaign analysis is the model developed in the analysis. The model should be presented in a 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 underpinning the analysis allows for better sensitivity analysis, both by the researcher and by potential critics or users. Readers with different information or beliefs about a parameter value can quickly assess how conclusions change as inputs change, if they have access to the model. Publishing the model in an easily reused form also 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.[104] They published the math behind their model, allowing researchers skeptical of any of their variable estimates to reimplement the model in code to see if their conclusion stood. In September 2019, eight years after Shifrinson and Priebe

---

101. Posen, *Inadvertent Escalation*, p. 68.
102. Mearsheimer, "Why the Soviets Can't Win Quickly in Central Europe," p. 3.
103. Cohen, "Toward Better Net Assessment," p. 57.
104. Shifrinson and Priebe, "A Crude Threat."

published their analysis, the availability of accurate cruise missiles enabled Iran's successful attack on the Abqaiq oil-processing facility in September 2019, just as their 2011 model suggested.

Second, 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 to build upon when studying other questions and scenarios of a similar type (e.g., air defense, nuclear counterforce strike, and logistical airlift operations). Some of the models used by researchers may become "workhorse" models, such as Richard Kugler's Attrition-FEBA (forward edge of battle area) Expansion model used by Posen and O'Hanlon, or Epstein's logistics model serving as the foundation for work by Ryan Baker.[105] 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: ADOPT 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 of each model variable. In most published campaign analyses, 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 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. In-

---

105. Posen, "Measuring the European Conventional Balance"; Posen, *Inadvertent Escalation*; O'Hanlon, "Stopping a North Korean Invasion"; Joshua M. Epstein, *Strategy and Force Planning: The Case of the Persian Gulf* (Washington, D.C.: Brookings Institution Press, 1987); and Baker, "Logistics and Military Power."

stead of running the analysis several times with different values to produce several different outcomes, researchers can use the input distribution approach to conduct sensitivity analysis on all model variables at once and to present the outcome as a single distribution of outcomes.[106] This allows researchers to calculate likely values, confidence intervals, or other statistics to convey the uncertainty of the model's output.

Modeling inputs as distributions cannot be done by hand, and requires computational tools. Monte Carlo techniques provide a simple way to implement this 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.[107]

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 the lethal radius equation above). 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.[108]

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

---

106. Previous studies have 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?"

107. As with many of the methods discussed in this article, Monte Carlo methods were developed during World War II. Herbert L. Anderson, "Metropolis, Monte Carlo, and the MANIAC," *Los Alamos Science*, Fall 1986, pp. 96–107, https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-86-2600-05. Some previous campaign analyses have used Monte Carlo techniques to show how the outcome varies as a single input takes on different values. Bell, "Can Britain Defend the Falklands?"; and Wu, "Living with Uncertainty." See also 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, https://www.jstor.org/stable/26553096.

108. Our approach has similarities to 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, for example, Richardson and Stone, "Operations Analysis during the Underwater Search for Scorpion."

tivity 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.

Representing the output as a 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 percent 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 uncertainty approach enables researchers to put more of the uncertainty directly into the model's output, rather than adding it on afterward.

Finally, the input distribution approach allows researchers to use more of the information they have already collected. The traditional approach to campaign analysis requires researchers to select and defend a single value for each input (or a most plausible and conservative value), when their research often suggests uncertainty around input values. Allowing researchers to specify ranges preserves the information they gather when conducting their research.

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 accounts only for uncertainty in inputs to the model, not for uncertainty about the model itself. If a model is misspecified, the actual outcome could be far outside the distribution it returns.[109] 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 understanding of the world.

Researchers also might be tempted to neglect the research they have

---

109. 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.

done and overcount 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 approach is advisable for these researchers.

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 Wu and Posen are exemplary in how transparently they document their models and parameter values. Second, the two analyses help 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 thirty-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.[110] 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. Finally, we use the replications to

---

110. Lieber and Press, "The End of MAD?"

demonstrate the value of campaign analysis for academic theory. The analyses provide alternative measures of two variables at the center of international relations debates: second-strike survivability and the ease of territorial conquest.

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.[111]

UNITED STATES–CHINA NUCLEAR COUNTERFORCE

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 U.S. attack.[112] Most recently, Wu has used campaign analysis to argue that the Chinese nuclear deterrent was far from assured at several points in China's past.[113]

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 U.S. 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."[114] By using one model and modifying model parameters to analyze different scenarios across two different dyads over a

---

111. Our tools are available at doi.org/10.7910/DVN/998QEK.
112. Charles L. Glaser and Steve Fetter, "Should the United States Reject MAD? Damage Limitation and U.S .Nuclear Strategy toward China," *International Security*, Vol. 41, No. 1 (Summer 2016), pp. 49–98, doi.org/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 (Fall 2019), pp. 61–109, doi.org/10.1162/isec_a_00359; Michael Chase and Evan Medeiros, "China's Evolving Nuclear Calculus: Modernization and Doctrinal Debate," *RAND/CNAC PLA Conference* (Santa Monica, Calif.: RAND Corporation, 2017); James C. Mulvenon et al., *Chinese Responses to U.S. Military Transformation and Implications for the Department of Defense* (Santa, Monica, Calif: 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 (Fall 2010), pp. 48–87, doi.org/10.1162/ISEC_a_00016; Fiona S. Cunningham and M. Taylor Fravel, "Assuring Assured Retaliation: China's Nuclear Posture and U.S.-China Strategic Stability," *International Security*, Vol. 40, No. 2 (Fall 2015), pp. 7–50, doi.org/10.1162/ISEC_a_00215; and Wu Riqiang, "Certainty of Uncertainty: Nuclear Strategy with Chinese Characteristics," *Journal of Strategic Studies*, Vol. 36, No. 4 (2013), pp. 579–614, doi.org/10.1080/01402390.2013.772510.
113. Wu, "Living with Uncertainty."
114. Ibid., p. 114.

twenty-five-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 U.S. 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 Wu 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, as other nuclear strike models do.[115]

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 U.S. 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 U.S.-China 2010 scenario, if the "criterion for deterrence" is a

---

115. Lieber and Press, "The End of MAD?"

single warhead surviving to retaliate, he finds a 38 percent probability of meeting the single-warhead threshold when nuclear forces are on day-to-day alert, and 90 percent probability when the missiles are on fully alerted status.[116] The probability that five or more warheads survive is 6 percent and 1 percent for full alert and day-to-day alert, respectively.[117]

ADAPTING A TRANSPARENT MODEL. We successfully replicate Wu's model of the U.S.-China 2010 scenario and reach the same probabilistic conclusions. We choose to replicate the U.S.-China 2010 scenario (one of eight scenarios Wu modeled), simply because Wu identified this as the "baseline for China-U.S. strategic stability,"[118] when mobile missiles first created a survivable deterrent.

We argue in recommendation 1 that models are often more generalizable than many researchers assume. We demonstrate the broader applicability of models, and thus the wider contributions a single campaign analysis can make, by reexamining Wu's 2010 U.S.-China counterforce attack using a slightly modified form of a previously published counterforce model by Press and Lieber.[119] Although the models are quite different, we find similar results using point estimates from Wu's article in Lieber and Press's model. Our findings with respect to Chinese nuclear survivability when we use Lieber and Press's 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's model estimates the number of warheads that are expected to survive a first strike, whereas 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 Chinese missile must survive at each step leading up to launch. Lieber and Press'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 percent for soft tar-

---

116. Wu, "Living with Uncertainty," online appendix 2 at doi.org/10.7910/DVN/5EKNJM.
117. Ibid.
118. Ibid., p. 87.
119. Lieber and Press, "The End of MAD?"

gets). 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 sought to replicate Wu's findings in the U.S.-China scenario by adapting Lieber and Press's existing counterforce model. We begin with Lieber and Press's model of nuclear combat, which involves the specific accuracy and yield of different U.S. 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 allowed 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's bolt-from-the-blue scenario. Second, we needed to include a term for detection probabilities in the model. Lieber and Press did not have 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 launchpads, forward bases, and technical sites as Wu's detailed, China-specific model does.

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

Our result for U.S.-China counterforce 2010 using Lieber and Press's 2006 U.S.-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 percent. If we use a more conservative detection probability of 0.70, the probability that a single warhead survives is higher than 90 percent, and there is a 50 percent 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, whereas 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 conclude that the Chinese arsenal is best described as "first-strike uncertainty." The attacker's ability to destroy all warheads with certainty is far from assured.

CONTRIBUTION TO THE ACADEMIC DEBATE—SECURE SECOND STRIKE. Campaign analyses such as the nuclear counterforce campaigns analyzed by Wu and Lieber and Press provide the most precise, publicly available measure of nuclear survivability, a variable at the core of academic debate on the nuclear revolution.[120] A central feature of the contemporary nuclear era is some loss of confidence in secure second strike given smaller arsenals and improved counterforce, particularly improvements in missile accuracy and technical intelligence.[121] 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?" Campaign analysis presents a significant improvement over proxy variables such as simple warhead tallies, which are merely one (and not necessarily the most significant) of the variables relevant to second-strike survivability.

(RE)ASSESSING THE CONVENTIONAL BALANCE IN EUROPE

In the 1980s, security studies researchers and policymakers debated the balance of conventional forces in Europe. Soviet numerical superiority led many observers 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. This work also opened up a methodological debate about how to analyze military scenarios, and the role of academics in doing so.

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, cam-

---

120. For seminal examples, see Robert Jervis, *The Meaning of the Nuclear Revolution: Statecraft and the Prospect of Armageddon* (Ithaca, N.Y.: Cornell University Press, 1989); Kenneth N. Waltz, "Nuclear Myths and Political Realities," *American Political Science Review*, Vol. 84, No. 3 (September 1990), pp. 730–745, doi.org/10.2307/1962764; and Stephen Van Evera, *Causes of War: Power and the Roots of Conflict* (Ithaca, N.Y.: Cornell University Press, 1999).
121. Lieber and Press, "The End of MAD?"; and 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 (Spring 2017), pp. 9–49, doi.org/10.1162/ISEC_a_00273.

paign-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 whether one side could achieve a 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 Kugler's "attrition-FEBA expansion" model. This model takes in several parameters, the most important of which are 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."[122] Using

---

122. Posen, *Inadvertent Escalation*, p. 68.

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."[123]

EMPLOYING THE INPUT DISTRIBUTION APPROACH. 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 in Posen's *Inadvertent Escalation*, which describe the formulas involved in calculating the Attrition-FEBA Expansion 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 employ 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 Monte Carlo techniques, we draw uniformly from these ranges and re-calculate 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 equally weight all of the values between the pessimistic conventional wisdom and the more NATO-favorable values Posen defends to determine whether Posen's conclusion is robust. Incorporating all of this uncertainty suggests that NATO forces face a non-zero chance of being overrun, but are likely to hold out. By the ninetieth day of the campaign, the probability of a NATO shortfall occurring at any point approaches 25 percent, although Pact forces would also be depleted here. 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 into an output that expresses NATO's competitiveness with the Warsaw Pact across a range of possible conditions. Under some of these combinations of variables, NATO forces

123. Ibid., p. 127.

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 THE ACADEMIC DEBATE: EASE OF CONQUEST. Campaign analyses such as Posen's can produce improved measures of the ease of territorial conquest, a variable at the center of offense-defense theory.[124] The offense-defense balance, like secure second strike, is both a key variable in academic international relations and very difficult to measure.[125] Campaign analysis improves upon alternative approaches to measuring the offense-defense balance by permitting researchers 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

124. Robert Jervis, "Cooperation under the Security Dilemma," *World Politics*, Vol. 30, No. 2 (January 1978), pp. 167–214, doi.org/10.2307/2009958; Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?"; and Van Evera, *Causes of War.*
125. Jack S. Levy, "The Offensive/Defensive Balance of Military Technology: A Theoretical and Historical Analysis," *International Studies Quarterly*, Vol. 28, No. 2 (June 1984), pp. 219–238, doi.org/10.2307/2600696; and Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?"
126. For a similar suggestion, see Glaser and Kaufmann, "What Is the Offense-Defense Balance and How Can We Measure It?" pp. 75–76.

India and Pakistan, or any other pair of potential adversaries, allowing re-searchers to test the predictions of the theory.

Conclusion

For decades, scholars have conducted campaign analysis to study military op-erations. Until now, however, methodological guidance for conducting and evaluating campaign analysis has been sparse. Recent work in security studies 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, a core method in the field. Standardizing campaign analysis, as with other methods, equips readers to more easily evaluate its use, enables a wider pool of re-searchers 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 sce-nario under investigation and can serve as the foundation for other research-ers' 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 mod-els as interactive calculators (as we do in the replications) for others 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 uncer-tainty through to the final outcome using Monte Carlo techniques helps ad-dress this concern.

This article 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 theoreti-cal debate. In the replications section of the article, we demonstrate how cam-paign analysis can produce measures for two variables at the core of canonical international relations debates: nuclear survivability and ease of conquest.

In the future, we see great potential for campaign analysis to answer an even wider set of questions. Campaign analysis could more often be used to exam-ine historical campaigns, helping us understand why an operation turned out the way it did, validating models, and revealing puzzles. Researchers also could 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 and facilitates focused attention on 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 political strategies such as sanctions could shape military outcomes or how alliance cohesion or fracture could shape conflict. Researchers could also examine the effects of the spread of biological weapons, climate change, or pandemics.

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 more often use existing models from the military modeling and simulations 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 decisionmaking, drawn from political science or using specific techniques such as wargames, tabletop 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 they are only meaningful if researchers choose 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. Scholars of security studies, 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 answering an optimality question. Albert Wohlstetter et al., "Selection and Use of Strategic Air Bases" (Santa Monica, Calif.: RAND Corporation, 1954).

128. See, for example, Hillestad, Bennett, and Moore, "Modeling for Campaign Analysis," p. 21.