9 Data processing for applications of dynamics-based models to forecasting[1]

Michael Gabbay, University of Washington

1. Introduction

The Forecast operational capability tracks and projects change along multiple dimensions in sociocultural entities and phenomena of interest. This chapter addresses issues related to representing a particular sociocultural system in a way that involves (a) only data available for an actual, ongoing situation that can be used to (b) reveal meaningful changes in the system over time and (c) implement models that can help users to anticipate those changes. These three elements distinguish data processing issues for forecasting from those in the Understand capability area, in which historical – not only current – data can aid in understanding sociocultural systems. They also separate forecasting from the "Detect" area, in which it suffices for algorithms to find patterns in data without predicting future behavior or supporting causal models of how the system evolves. This chapter centers on selected aspects of data processing relevant to dynamics-based models, i.e., models that posit causal mechanisms of system evolution, rather than purely statistical models, and illustrates the discussion with applications in the domain of political and insurgent network dynamics.

The first section of this chapter briefly describes dynamics-based modeling methodologies and presents a model of group decision making as an example that will help guide the subsequent discussion. Dynamics-based models involve an important distinction between variables and parameters and the second section of the chapter discusses the requirements and considerations involved in calculating their values. Calculating parameters is particularly difficult and the extent to which they can be determined influences whether the output of a model can be evaluated with respect to an actual time scale or only with respect to equilibrium outcomes.

The chapter then presents two examples of applications, both concerning leadership elites and their organizations. Research has largely overlooked the behavior of leaders and organizations amidst the surge of interest in social media, population sentiment, and big data, yet this type of data is particularly important in tracking and anticipating change on time scales of operational concern with regard to entities subject to operationally relevant levers of influence. These two examples, presented in the fourth and fifth sections, use different types of input data. The first involves the use of expert judgments in conjunction with the group decision-making model. The second involves the representation and modeling of insurgent network behavior and shows how

---

[1] Approved for Public Release; Distribution Unlimited. Case Number 14-2487
This work was supported by Department of Defense Contract N00014-13-1-0381 and HDTRA1-10-1-0075

the ideologies, policies, and relationships among insurgent groups can be constructed using rhetoric as data and tracked over time. For the insurgent network application, the chapter presents a stochastic model of insurgent cooperation at the tactical level and compares the model results with the network's evolution as observed in the data. Finally, the chapter provides a brief discussion of gaps in knowledge and future research directions related to data processing and forecasting, focusing on areas relevant to dynamics-based models and political application domains.

2. Dynamics-Based Models

In dynamics-based models, variables can evolve based on their present state and their interactions with other variables. Thus, a variable can be affected by self-feedback, as in an exponential growth or decay situation in which the rate of change of the variable is proportional to its value, and by feedback from other variables in the system, as in a predator-prey model where the rate at which the number of prey animals declines depends on the number of predators. Typically, these models posit causal mechanisms that underlie the forms of the feedbacks included. For example, in a model that treats the spread of a revolutionary ideology like the spread of an epidemic, the rate at which the number of people "infected" with the ideology increases is taken to be the product of the present number of already infected people and the number of people who are susceptible but not yet infected (Epstein, 1997). This assumes that each interaction between an infected and susceptible person has a certain chance of resulting in a new adherent to the ideology, as in person-to-person disease transmission.

This product interaction between variables is an example of a nonlinear interaction. Nonlinear models can accommodate complex behaviors such as abrupt, discontinuous changes in the nature of the system's behavior. Such sharp transitions are often of highest importance in sociocultural forecasting, e.g., will a localized uprising result in a mass revolution? Indeed, the epidemic model contains a critical threshold, called the reproduction rate, above which the epidemic or revolution will take off but below which it will fizzle out. Purely statistically-based approaches such as regressions are ill suited to forecasting sudden transitions in sociocultural contexts, which is one reason why this chapter focuses on dynamics-based models. In addition, statistically based models often provide little insight into how interventions into a situation will play out; therefore, they cannot be applied to analyzing courses of action for the purposes of the Mitigate capability area. However, regardless of whether one uses dynamics-based models or statistical methods, sociocultural forecasting remains a difficult task, largely due to the lack of appropriate empirical data.

There are various ways of implementing dynamics-based models, three of which are briefly mentioned here: differential equations, stochastic models, and agent-based models (for a more detailed description of various modeling methods, see chapter 10 on computational models in the Forecasting section of this book). Differential equations give mathematical expressions for the rates of change in a system of variables that can then be solved, computationally or analytically, to show the values of the variables over time. The equations evolve variables in continuous time (difference equations are the discrete-time analog). The epidemic model noted above is most readily expressed and analyzed in differential equation form. The Lanchester model for combat dynamics is another differential equation that may be familiar to some readers due to its prevalence in the combat simulations used in operations research (Epstein, 1997). Differential equations are typically deterministic in that the future evolution of the variables is precisely determined by their initial values (although they can be adapted to accommodate stochastic forcings). A class of critical transitions in which the state of the system can change suddenly and discontinuously, known as “bifurcations,” can occur in nonlinear differential equations; for example, old states become unstable and disallowed while new stable states emerge.

Unlike differential equations, stochastic models allow variables to evolve in a random rather than a deterministic manner. Also, these models typically represent temporal evolution discretely rather than continuously. Often the models assume a Markov process in which the probabilities of the various future states of the system depend only on its current state. In a stochastic simulation of the epidemic model of ideology diffusion, one could simulate all the individuals in a population, endow them with different characteristics, and allow infected individuals to meet and infect susceptible individuals with given probabilities. Stochastic models are often used in simulating processes that occur over networks, such as opinion change and cultural transmission, and in simulating network evolution itself, as in the model of insurgent cooperation presented below. Scale-free networks – characterized by power law distributions in which a relatively few nodes have a greatly disproportionate degree (number of ties) – can be generated by a stochastic “preferential attachment” model in which new nodes have a higher probability of linking to high-degree nodes than to low-degree ones (Newman, 2010). Stochastic models can exhibit critical transitions resulting in abrupt changes of state in the form of phase transitions at large population sizes.

At its most general, agent-based modeling involves the specification of interaction rules for a system of agents in a group or population. These rules can be deterministic or probabilistic and expressed in terms of mathematical formulas, similar to differential equations and stochastic models. However, agent-based modeling diverges most notably from these other modeling approaches when the rules are essentially algorithmic: specifically, agent behavior is governed by a relatively complex process that accounts for strategies, heuristic rules, experience from prior interactions, and learning. Consequently, agent-based models can simulate very rich behavior, although the downside for forecasting purposes is that they usually require more data to implement empirically and more computational resources to simulate the parameter space for assessing potential outcomes. Examples of applications of agent-based models to social systems in specific real-world cases include the modeling of social identity dynamics (Alcorn, Garces, & Hicken, 2012) and political party competition (Laver & Sergenti, 2012).

Group Decision Making Model

To provide a specific example of a dynamics-based model and a more concrete basis for the discussion of empirical implementation issues, we now consider a nonlinear model of group decision making – one that has been applied for forecasting purposes (Gabbay, 2007; 2013). The model captures the evolution of group member policy positions under the influence of group discussion and their own ideological preferences. Although the model is formulated as a differential equation system, it is expressed here only in terms of qualitative rules to avoid becoming bogged down in mathematical detail. The rules are:

1. Each group member's current position is affected by two forces: the *self-bias force* and the *group influence force*.
2. The self-bias force is the tension that an individual feels when his current position is different from his *natural bias* position, which is determined by ideological and strategic preferences such as the relative importance placed upon military, diplomatic, economic, and political factors. The self-bias force:
   a) Acts in the direction of the natural bias.
   b) Has magnitude proportional to the difference between the individual's current position and natural bias.
   c) Increases in proportion to the individual's *commitment* to his natural bias.
3. The group influence force is the tension that an individual group member feels when his position is different from those of other group members.
   a) The total group influence force on an individual is the sum of the coupling forces resulting from his pairings with all the other group members.
   b) For a given pairing, the coupling force on an individual acts in the direction of the position of the other member of the pair.
   c) The magnitude of the coupling force increases approximately linearly for small position differences but weakens for differences greater than the individual's latitude of acceptance.
   d) The coupling force that member *j* exerts on member i is scaled by the *coupling strength*, which characterizes factors such as how often *j* communicates with *i* about the policy matter at hand, their relative status, *i*'s perception of *j*'s credibility or expertise, and the importance that *i* attaches to group influence.

At high disagreement levels (i.e., large differences in natural bias) the model displays nonlinear behaviors, including sharp transitions between qualitatively distinct outcomes as the coupling strength increases; asymmetric, majority rule outcomes resulting from symmetric conditions; the existence of multiple stable outcomes for the same parameters; and the greater facility of less dense networks to reach decisions and reduce discord. These behaviors are not present at low disagreement levels.

3. Variables and Parameters

All dynamics-based models are characterized by variables and parameters. Variables define the state of the system as it evolves in the model over time. In the model above, the policy positions of the individual group members are the state variables. Other examples of variables in sociocultural models include cultural norms, attitudes, popular support levels for government or insurgents, sizes of social movements, violence levels, and network ties among individuals or organizations. For forecasting applications, the state variables should be either: (1) indirectly related to the ultimate object of interest, in the sense that the application is intended to predict and measure an aggregated function of the state variables; or (2) the objects of ultimate interest themselves, in that the application is intended to predict and measure the future values of the state variables. For the group decision-making model, the first case would apply if the researcher's ultimate goal were only to forecast the final policy decision of the group. This can be done by aggregating the final policy positions of the individual group members according to a suitable decision rule (e.g., leader choice, weighted averaging, consensus). The second case would apply if the researcher also wanted to predict which group members will support the policy and which ones will dissent, which can be determined from the distance between their final position and the final policy. Another example of aggregating variables, as in (1) above, is a model of insurgency in which the state variables are the attitudes of individual population members toward the government and toward the insurgents, but the application seeks to forecast only the overall levels of support.

Parameters are model quantities that characterize the system exogenously, i.e., they are not affected by the state variables. Parameters play a key role in dynamics-based models in determining how variables interact with each other and the environment. In the epidemic model of militant ideology spread, parameters determine the ease with which the ideology can spread between individuals and the rate at which infected individuals are removed by government repression or counterinfluence. Referring to the group decision-making model above, the natural bias, is used to set the initial positions of the members; the commitment scales the strength of the self-bias force for a given displacement of each member's policy position from her natural bias; a person's latitude of acceptance sets the range of policies surrounding her own position that she will entertain and serves as the source of the nonlinear interactions; and the coupling strength between individuals scales the impact of relational factors on how effective a person will be at swaying another person's position. For stochastic models, the parameters determine how the transition probabilities between states depend upon the state variables.

Modelers usually take parameters as constant, but can make them time-dependent as well. Given the complexity of sociocultural systems, variables often provide feedback to parameters, but as long as the time scale of that feedback is slow relative to the time scale at which variables evolve one can still consider the parameters as effectively exogenous. If that feedback occurs at a rate comparable to the change in the variables, however, then the modeler faces a coevolution problem and should redefine the "parameters" as state variables. The group decision-making model example assumes that the members' changing policy positions have no impact on the coupling network between group members. This is reasonable for a relatively short decision-making episode, but if the model were examining a long time scale—say years—then policy positions would probably affect relationships under the operation of the homophily principle: "birds of a feather flock together."

One of the primary difficulties – perhaps the greatest one – in forecasting sociocultural phenomena is that many of the concepts used in social science theory as well as in common discourse are not readily quantifiable. Power, ideologies, policies, attitudes, beliefs, grievances, trust, relationships, disagreements, utility, etc., do not come in standard units like watts, meters, or amperes. For instance, a social network matrix may use a single number to represent ties between individuals, but that does not capture the complexity of social relationships.

Forecast, therefore, occupies a thorny middle ground in the abstract-to-concrete data spectrum between the Understand and Detect capabilities. To help *understand* general decision-making phenomena, one can model policy spaces in the abstract, but one must assign actual policies to policy space coordinates in order to *forecast* whether a particular country's leaders may decide upon war in a given situation. For *detection*, on the other hand, very specific, concrete, and readily quantified data such as the frequency of conversations may suffice to identify central individuals or communities in, say, an email network. However, that is not equivalent to identifying the top leaders who make strategic decisions in times of crisis and the relationships of authority and influence between them. Consequently, a key challenge in forecasting is to construct meaningful variables and parameters from relatively specific, more quantifiable, lower level data that can be reasonably mapped onto more abstract, higher order theoretical concepts.

This construction of conceptually broader variables and parameters is necessary not simply to connect with social science theory but also to reduce the dimensionality of the space of variables that must be predicted. Considering the group decision-making context, a group member's natural bias policy may be a function of her beliefs about the facts, the relative importance that she accords to the different components of the problem – military, diplomatic, economic, domestic political – and how she evaluates the policy options for each of those components. Although an analyst may be able to assess those factors for each group member, no model can be expected to predict the changes in each of those factors due to group interactions. It is therefore necessary to reduce the dimensionality of the space of the state variables by, for instance, combining the above factors to represent policy positions in a one or two-dimensional space. The subsections on analyst input and insurgent rhetoric present examples of ways to construct ideology and policies from lower level data.

Variables and some kinds of parameters can be calculated by applying the defining formula or algorithm to data representing a single time instant or interval independently of other times or other variables. However, feedback parameters in particular, which necessarily involve rates of change over time or the strength of interaction among variables, cannot truly be calculated solely on the basis of a predefined formula or algorithm that operates on data representing only one time. Ideally, one would measure the behavior of the state variables over time and then fit the parameters to the data, similar to the way parameters are measured experimentally in physical systems. This is also analogous to statistically based pattern recognition and forecasting methods in which an in-sample is used to fit parameters that are then applied to out-sample data. Such a fitting procedure can be used to determine parameters in sociocultural systems, but typically is difficult to carry out to much precision due to the lack of sufficient data points over time and the relatively large amount of noise in the data, among other factors. This is particularly true of the conflict-ridden and chaotic regions of interest to the U.S. military. When fitting parameters to existing data is not an option, then modelers can choose parameter values in an a priori manner based on considerations of what constitutes reasonable values or ranges. Simulations can then sweep over a patch of parameter space in order to determine the sensitivity of the outcomes to changes in the parameters. Laver and Sergenti (2012) offer a thorough and candid discussion of the issues involved in parameter calculation for an agent-based model of political party competition.

4. Forecasting Modes

Dynamics-based models can be applied for forecasting purposes in a variety of modes. The modes available for a given application depend on the nature of the data. One mode involves predicting, in a probabilistic way, the evolution of variables that can change in a more-or-less continuous fashion over time: for instance, the levels of popular support for an opposition movement among different segments of the population. This would correspond to true forecasting in that one assigns time-dependent probabilities of the various possible states of the system, as is done in weather forecasting. However, as discussed above, this remains an ambitious goal for most models of sociocultural dynamics due to the limitations in measuring variables and parameters (as well as model uncertainty itself, a topic not considered in this chapter). Although it would be technically possible to develop simulation software that makes explicitly probabilistic forecasts, doing so could well be counterproductive in the absence of accurate and precise estimates of model variables and parameters. Otherwise, end users, accustomed to reliable weather forecasts, could be lulled into a false sense of confidence by the simulation's provision of numerical probabilities and, consequently, might be greatly disappointed by the inevitable large mismatch between the forecasts and reality.

A looser mode of forecasting can be applied to key events, such as conflicts, decisions, and alliance composition, which can be treated as one-off occurrences. In these instances, one relaxes the requirement that the model make forecasts over time and simply assumes that the time window is sufficiently long to allow the system to reach equilibrium. In the group decision-making model, the time scales that enter into the commitment and coupling scale parameters are not usually known, which precludes such a model from being used to forecast the evolution of the group members' policy positions in real units of time. However, the equilibrium positions themselves do not depend on the time scales, and so the model can still be used to predict the final decision and its supporters and dissenters, on the assumption that the debate persists until equilibrium. This equilibrium approach significantly eases data demands; many dynamics-based models use this mode in practice.

Sensitivity analyses in which simulations cover a region in the parameter space can also be employed for forecasting purposes. If the outcome does not vary significantly as parameters sweep over a range of reasonable or likely values, then the model user can conclude that the outcome is highly probable. Alternatively, the simulations may reveal a range of possible outcomes. These may include potential outcomes (e.g., alliance configurations) not anticipated prior to the simulation. Sensitivity analysis may also reveal that the outcome can change dramatically with only a small shift in parameter values, as would be consistent with a bifurcation or phase transition.

As an illustration, consider a group decision-making model in which all group members are connected with the same coupling strengths and have initial opinions (natural biases) symmetrically distributed around zero. More specifically, consider a three-person group in which a centrist is bracketed by two equidistant extremists. Standard intuition would anticipate either a deadlock or various shades of compromise around the centrist position, consistent with final states that are symmetric about the middle, as shown in Figure 1 (a) and (b) for deadlock and compromise respectively. However, at sufficiently high levels of initial disagreement, another outcome can result in which the centrist swings toward one of the extremes (depending on random perturbations), corresponding to a majority rule situation favoring one side of the policy axis. In this case the system reaches an asymmetric final state as observed in Figure 1(c) (Gabbay & Das, 2014). Even a small change in coupling strength can produce a transition to the majority rule outcome zone from either the deadlock or compromise zones. If the simulation of a leadership group showed such behavior, analysts could be prepared for an extreme decision rather than deadlock or compromise.

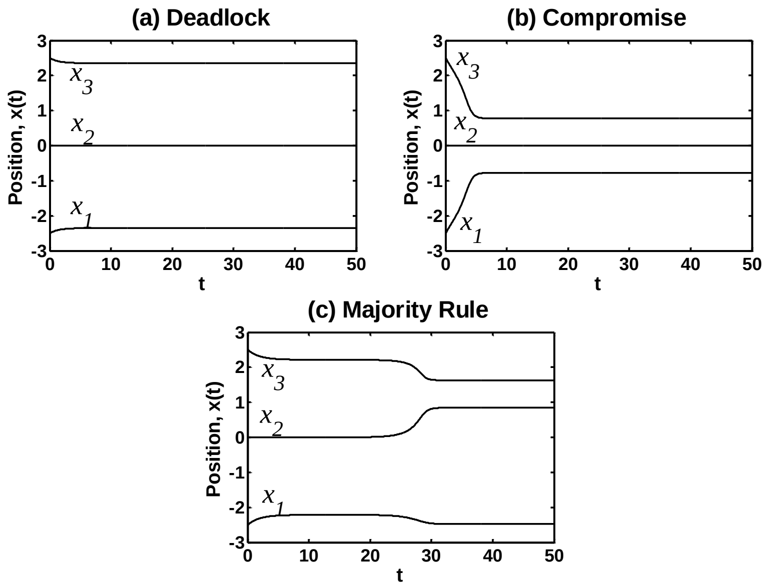

Figure 1. Group decision-making simulation of member positions over time in triad network (chain topology) with symmetric coupling strengths and natural bias distributions: (a) symmetric deadlock outcome at low coupling strengths; (b) symmetric compromise outcome at high coupling strengths; and (c) asymmetric majority rule outcome at intermediate coupling strengths. Adapted with permission from Gabbay & Das (2014).

Finally, dynamics-based models can be used to conduct scenario analyses to forecast the effects of specific hypothesized contingencies or conditions. These analyses can be implemented by particular variable or parameter settings that represent exogenous impacts on the system due to events or actions outside the dynamics of the model. For example, in a group decision-making situation, one can simulate the effect upon the policy outcome of a rupture in the relationship between two group members due to a personal dispute unrelated to the policy context. This could be done by nulling or severely attenuating their mutual coupling strengths, possibly in combination with other parameter changes.

5. Analyst Input for Group Decision Making

The two principal sources of data for calculating the "soft" variables and parameters (opinions, beliefs, ideologies, norms, social identities, etc.) needed as model input are analyst (subject matter expert) judgments and content analysis. Currently, analyst-based input represents the prevalent form of input (e.g., Alcorn et al., 2012; Feder, 2002; Gabbay, 2013), but the use of content analysis-based input is growing, especially given recent advances in automated text analysis. This section briefly discusses the use of analyst input in conjunction with the group decision-making model described above. The next section describes an application of content analysis using rhetoric as input data for the context of insurgent network dynamics.

Figure 2 shows the processing chain to implement the model for specific leadership groups using analyst judgment for input. The modeler obtains this input from a survey given to one or, preferably, many analysts. A composite analyst can be formed by averaging the survey responses of the individual analysts. The aggregation of individual surveys allows analyst judgments to be synthesized independently of each other. This minimizes the chance that social pressure will alter individual judgment, as can happen if the modeler elicits inputs in an oral discussion with a group of analysts – a common practice in other models of group decision-making used within the national security community (Bueno de Mesquita, 2009). Note that results can be generated on the basis of individual surveys as well. This permits comparison of the results obtained from individual analysts with the composite analyst and with each other, thereby providing a way of stimulating debate about differences between analyst viewpoints.

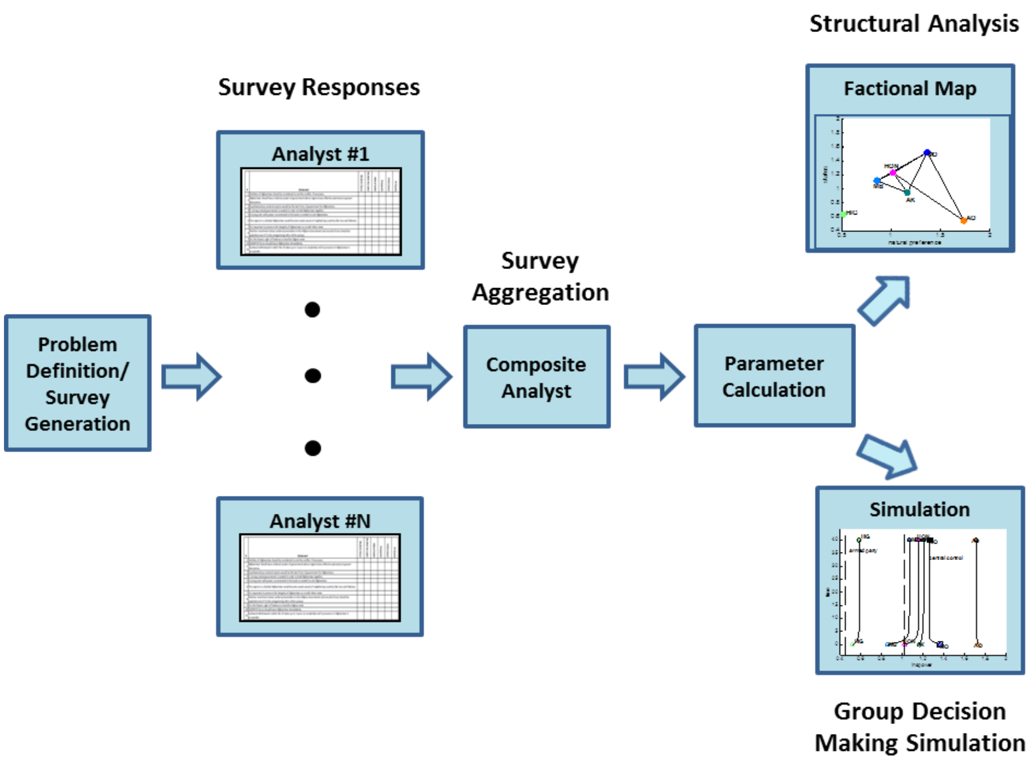

Figure 2. Overview of analyst input-based implementation of group decision-making model. Adapted with permission from Gabbay (2013). Copyright held by the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering.

Another innovation in this implementation methodology is the use of an attitude scaling technique to assess member ideological and policy positions. The natural bias, which is typically taken as the initial position of a group member, is constructed on the basis of a series of attitude statements for which the analyst rates the level of agreement or disagreement as if she were the group member in question, as shown in Table 1. This enables assessment of several different facets of the policy problem that can then be aggregated to calculate the natural bias. Also, modelers can probe the deeper structure of the policy space by using matrix decomposition methods such as principal component analysis. This approach elicits analysts' expertise on group member policy preferences without demanding that they directly perform the abstraction needed to create a policy axis or space itself — a task that the modeler accomplishes instead.

Table 1. *Sample attitude statements assessing ideological and policy preferences from an analyst survey on Afghan government elites. Analysts are asked to rate member attitudes on a five-point scale ranging from "strongly disagree" to "strongly agree"*

#	**Statement**	**Karzai**	**Fahim**	**Khalili**	**Rabbani**	**Dostum**	**Nur**
1	Partition of Afghanistan should be considered to end the conflict, if necessary.						
2	Afghanistan should have a federal system of government where regions have effective autonomy to govern themselves.						
3	Karzai's efforts to concentrate power in the presidency show that the Afghan Constitution should be changed to institute a parliamentary -centered system of government.						
4	A strong central government is needed in order to hold Afghanistan together.						
5	A strong leader with power concentrated in his hands is needed to rule Afghanistan.						
6	The regions in a divided Afghanistan would become weak vassals of neighboring countries like Iran and Pakistan.						
7	It is important to preserve the integrity of Afghanistan as a multi-ethnic state.						
8	It is the historic right of Pashtuns to lead the Afghan state.						

6. Rhetoric Input for Insurgent Network Dynamics

This section presents an application using input derived from content analysis – both automated and manual – in which the data source is Iraqi insurgent rhetoric. It describes methods for quantifying insurgent ideology, targeting policy, and cooperative relationships. Those three elements roughly correspond to insurgent ends, means, and allies respectively: critical components of their strategic behavior and decision making. These elements can be used to track insurgent behavior over time and as inputs in models for forecasting insurgent dynamics. A model of insurgent tactical cooperation is presented at the end of this section as an example of how such data can be used for simulating insurgent network evolution.

The modelers collected data for the 18 Iraqi Sunni insurgent groups listed in Table 2, spanning the time from mid-2003 through mid-2009. This time span is divided into three periods: (1) Period 1, August 2003 – July 2005; (2) Period 2, August 2005 – July 2007; and (3) Period 3, August 2007 – July 2009. The data set consists of roughly 2,000 translated insurgent statements from jihadist websites and interviews of insurgent group officials in print and broadcast media as provided by the U.S. government's Open Source Center.

Table 2. *Sunni insurgent groups in Iraq used in content analysis of insurgent rhetoric*

Group	**Symbol**	**Overall Classification**	**Islamist Ideology**	**Time Periods**
Al-Qaida in Iraq	AQI	Jihadist Salafist	Salafist	1,2,3
Ansar al-Sunnah Army (Ansar al-Islam: post-Dec. 2007)	ASA	Jihadist Salafist	Salafist	1,2,3
Islamic Army in Iraq	IAI	Nationalist	Salafist	1,2,3
Mujahidin Army	MA	Nationalist	Salafist	1,2,3
1920 Revolution Brigades	1920RB	Nationalist	Unspecified	1,2,3
Islamic Front for Iraqi Resistance	JAMI	Nationalist	Muslim Brotherhood	1,2,3
Rashidin Army	RA	Nationalist	Unspecified	1,2,3
HAMAS-Iraq	HAMI	Nationalist	Muslim Brotherhood	3
Fatihin Army	FA	Nationalist	Salafist	2,3
Iraqi Jihadist Leagues	IJL	Jihadist Salafist	Salafist	2,3
Shield of Islam Brigade	SIB	Jihadist Salafist	Salafist	2,3
Ansar al-Sunnah Shariah Commission	ASA-SH	Nationalist	Salafist	3
Just Punishment Brigades	JPB	Jihadist Salafist	Salafist	2,3 (part)
Abu Bakr al-Siddiq Salafi Army	ABSSA	Jihadist Salafist	Salafist	2,3
Islamic Jihad Brigades	ISJIBR	Jihadist Salafist	Salafist	1,2
Victorious Sect Army	VSA	Jihadist Salafist	Salafist	1,2
Saad Bin Abi Waqqas Brigades	SBAW	Jihadist Salafist	Salafist	2,3
Army of Naqshabandi Order	NAQSH	Nationalist (Baathist)	Sufi	2,3

Ideology

Ideology is quantified by the concept of a *conflict frame*. The conflict frame of a political actor, such as an insurgent group or government leaders, is defined on the basis of the set of out-group parties perceived as current or potential enemies and in-group parties perceived as allies or as a base of support (Gabbay & Thirkill-Mackelprang, 2011). This definition stems from social identity theory and its implications for political rhetoric. *Conflict parties* consist of broad groups such as the incumbent government, ethnic or religious groups and their leadership classes, and foreign states. The mathematical formalism for conflict frames relies on the frequency with which specified marker terms appear in actor rhetoric and their in/out-group valences. Figure 3 is a conceptual diagram of the elements involved in the conflict frame calculation procedure. Sample marker terms and associated in- and out-groups are given in Table 3.

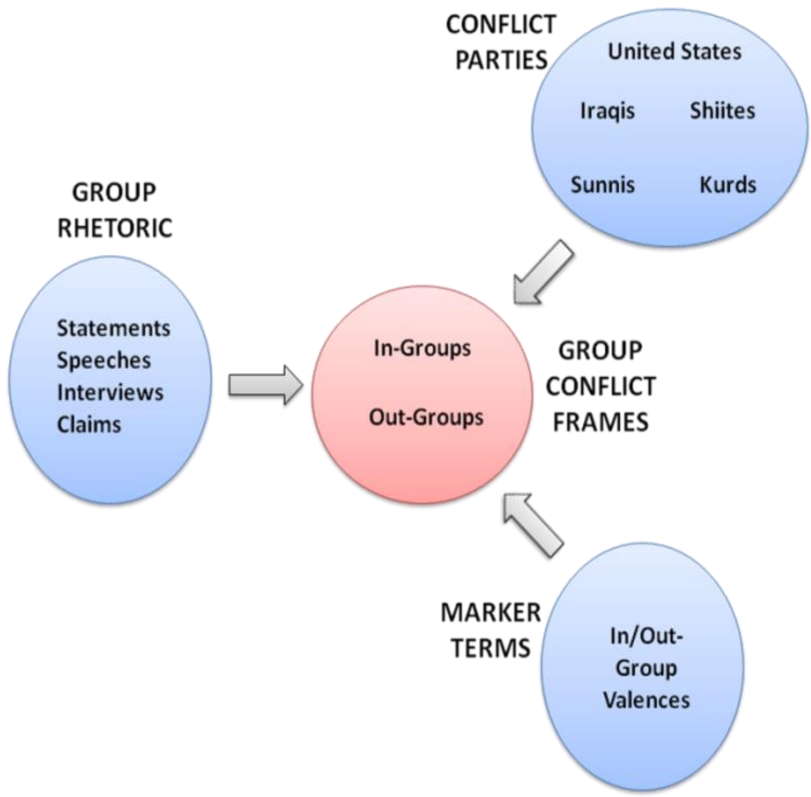

Figure 3. Conceptual diagram of conflict frame method of quantifying ideology.

Table 3. *Examples of marker terms and associated in-groups and out-groups for Iraqi insurgents*

Marker Term	**In-Groups**	**Out-Groups**
agent government		Iraqi Government
apostate government	Sunni Civil Society	Iraqi Government
companions of the prophet	Sunni Civil Society	
cross worshippers	Sunni Civil Society	United States
Iraqis	Iraqi Civil Society	
Iranian occupation	Sunni Civil Society	Iraqi Government, Shiite Political Parties
occupier		United States
rejectionist	Sunni Civil Society	Shiite Civil Society
rescue council		Sunni militias (Awakening Councils)

Each marker term is assigned a valence for all the conflict parties to which it refers: positive (+1) for in-groups, negative (–1) for out-groups, and zero for neutral references. A conflict party's *salience* to the actor is essentially the frequency with which marker terms referring to that party appear in the actor's rhetoric, regardless of valence. The actor's *attitude* toward the party is related to the relative frequency of positive and negative references to it. The attitude weighted by the salience represents the actor's *orientation* toward the party. An actor can have a highly negative attitude toward a conflict party but a low orientation value if the actor seldom refers to the party, which hence has low salience. The collection of orientation values for all the conflict parties forms the conflict frame of the insurgent group.

This procedure is automated following the construction of the initial dictionary and valence matrix and therefore can be updated in near-real time to track ideological shifts and divisions within the actor ensemble. An actor's conflict frame vector, composed of the actor's orientations toward the conflict parties, is the primary output of the algorithm. Another way of analyzing the data, however, is by considering certain master frames, which can be given intuitive interpretations. A master frame consists of a specified subset of out-groups and in-groups. The extent to which an actor espouses a given master frame can be quantitatively gauged by how closely its conflict frame aligns with the master frame – mathematically via the inner product of the corresponding vectors. Two master frames dominate for the Iraqi insurgents: (a) the resistance frame that pits Iraqis as in-group against the United States as out-group and (b) the sectarian frame that pits Sunnis as in-group against Shiites as out-group.

Figure 4 shows the conflict frames for the Iraqi insurgent groups over the three time periods.[2] The Jihadist groups—particularly the two major ones, Al Qaida in Iraq (AQI) and Ansar al-Sunnah—tend to fall on the high end of the sectarian frame and the low end of the resistance frame. The Nationalists, especially "pure" Nationalist groups such as the 1920 Revolution Brigades, JAMI, and the Rashidin Army, have lower sectarian frame values and high resistance values. However, it is important to note that the figures show a spectrum rather than two widely disjoint clusters of Jihadists and Nationalists, as would be implied by such a binary analytical classification. This spectrum quality and the locations of individual groups in the master frame coordinate space have meaningful implications for cooperative behavior among insurgent groups. The first two Nationalist alliances to nucleate comprised neighbors in the Period 2 master frame space: (1) the Jihad and Reform Front consisting of IAI, MA, FA, and ASA-SH; and (2) the Jihad and Change Front consisting of 1920RB and RA. This suggests that the conflict frame construct can provide meaningful and predictive input to models of insurgent alliance formation.

[2]The groups ABSSA and ISJIBR are not included (documents were not preprocessed to remove extraneous text).

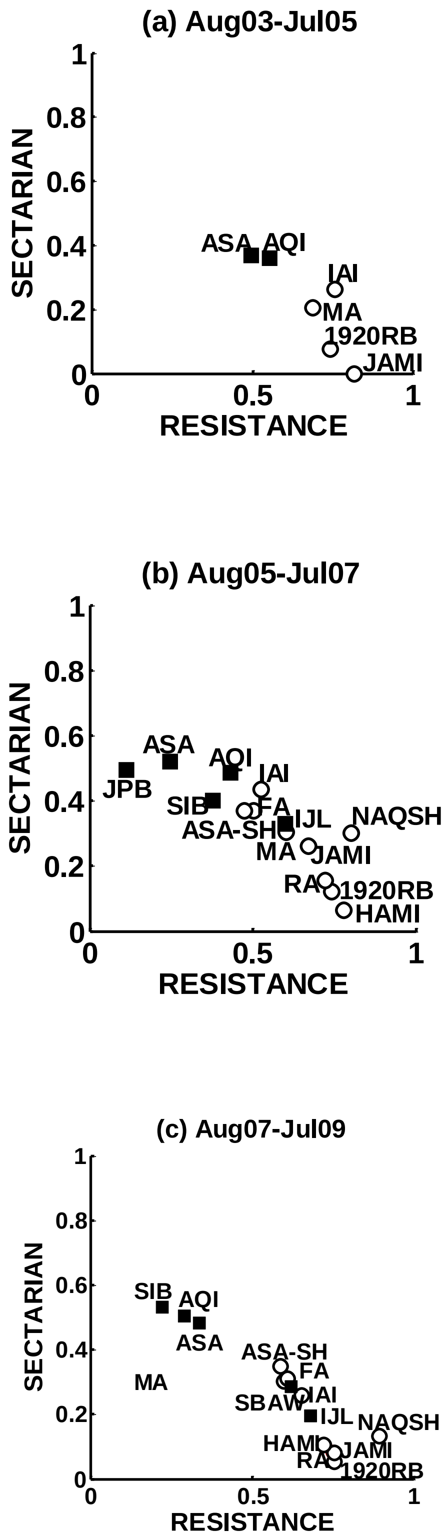

*Figure 4***.** Master frames for Iraqi insurgent groups: (a) Aug03–Jul05; (b) Aug05–Jul07; (c) Aug07–Jul09. Jihadists in black squares, Nationalists in white circles.

Tracking changes is one of the objectives of the Forecast area and the time-dependent behavior of the conflict frames can be tracked for either individual groups (as in Figure 4) or the whole ensemble. For example, in tracking changes in the balance between the two master frames for whole ensemble of groups, Figure 5(a) shows that the sectarian frame component increases relative to the resistance frame from the first to the second period, even among the Nationalists, and then wanes again in the third. This tracks the transition into a Sunni-Shiite civil war during 2006 and 2007 and the subsequent decline in sectarian violence. Ideological dissension within the insurgency can be tracked by looking at the frame deviation: i.e., how spread out the groups are in the space of master frames. Figure 5(b) shows the frame deviation in the resistance-sectarian master frame space. The deviation starts out low during the first period but increases and plateaus for the second and third periods. This corresponds to the character of factional relations within the insurgency. Early on, the groups were relatively united, but the transition to a sectarian civil war was a key factor in producing a rift between the Jihadist and Nationalist wings of the insurgency, eventually leading to the formation of the anti-AQI Awakening Councils and open fighting between Nationalist groups and AQI.

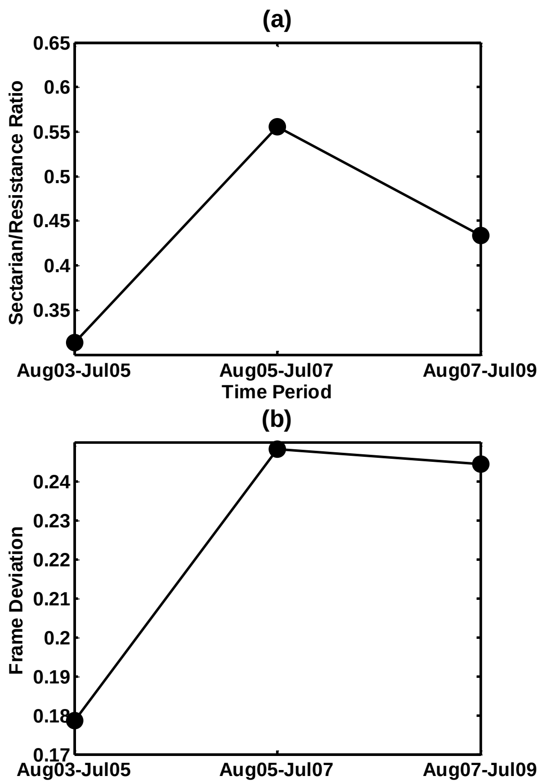

Figure 5. Temporal tracking measures for Iraqi insurgent ideology: (a) Ratio of mean sectarian frame value to mean resistance frame value; (b) Standard deviation in group two-dimensional frame positions from mean position of sectarian and resistance frames.

Targeting Policy

To construct the targeting policy variable, the portfolio of target classes—U.S. troops, Iraqi security forces, Shiite militias, government officials, civilians, etc.—claimed by different insurgent groups is considered (Gabbay & Thirkill-Mackelprang, 2011). The targeting policy scores each insurgent group by the average legitimacy of its target class portfolio, where the "legitimacy" of each target class is the acceptability of attacking it as determined by the prevalence, within the set of insurgent groups, of claims and statements supporting targeting the class vs. condemnations of doing so. A high targeting policy corresponds to a more discriminate or selective use of violence, while a low one indicates more expansive and controversial targeting practices. Implicit in this construction is that distance along the targeting policy axis is related to the extent of disagreement. This reflects that disagreement over the legitimacy of different types of targets has often been the primary source of dissension within Islamist insurgencies (Hafez, 2003).

Figure 6 shows insurgent group targeting policies over the second and third time periods. The groups are arranged along the vertical axis in ascending targeting policy order so that groups at the bottom have lower targeting policies. The figure shows that the Jihadists tend to have lower targeting policies and Nationalists higher ones. However, similar to conflict frames, this reflects a spectrum rather than a binary distribution. Like conflict frames, targeting policy can also be used to track dissension among insurgent factions; the standard deviation of the targeting policy increases over time in tandem with the actual rise in dissension within the insurgency (Gabbay & Thirkill-Mackelprang, 2011).

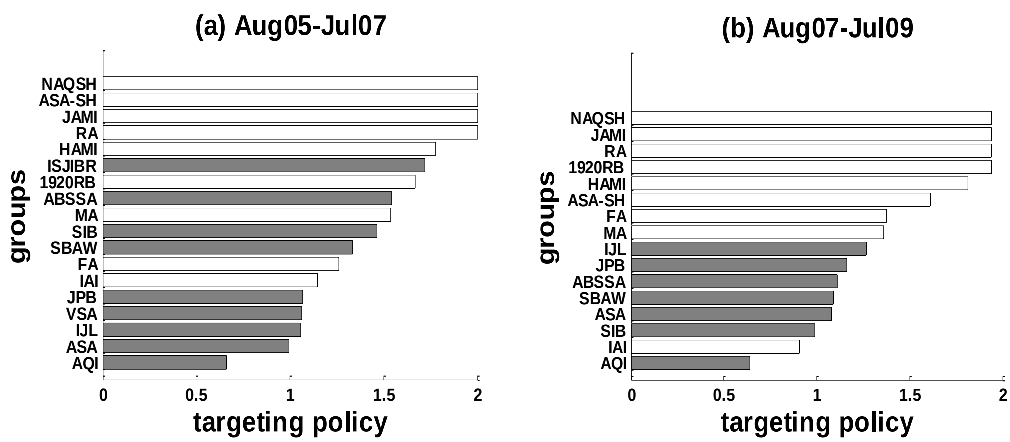

Figure 6. Targeting policies for Iraqi insurgent groups: (a) Aug05 – Jul07; (b) Aug07 – Jul09. Jihadists in gray, Nationalists in white.

Cooperative Networks

The rhetoric-based methodology constructs networks representing cooperative relationships among insurgents at both the leadership and rank-and-file levels. Leadership relationships among groups are gauged by the number of joint communiqués they issue. A joint communiqué is a statement signed by two or more groups, indicating the presence of communication and some level of agreement among the leadership of the groups issuing it. Furthermore, it demonstrates a willingness of the groups involved to be publicly associated with each other. At the rank-and-file level, the method uses the number of joint operations between groups. Typically, only one of the participants makes a claim of joint operations. Such operations indicate tactical coordination among insurgent groups, although presumably a group's leadership would have to approve the public disclosure of such cooperation. Figure 77 (a) shows the joint operations network for the third time period; Figure 77 (b) shows the simulated version, which is discussed presently.

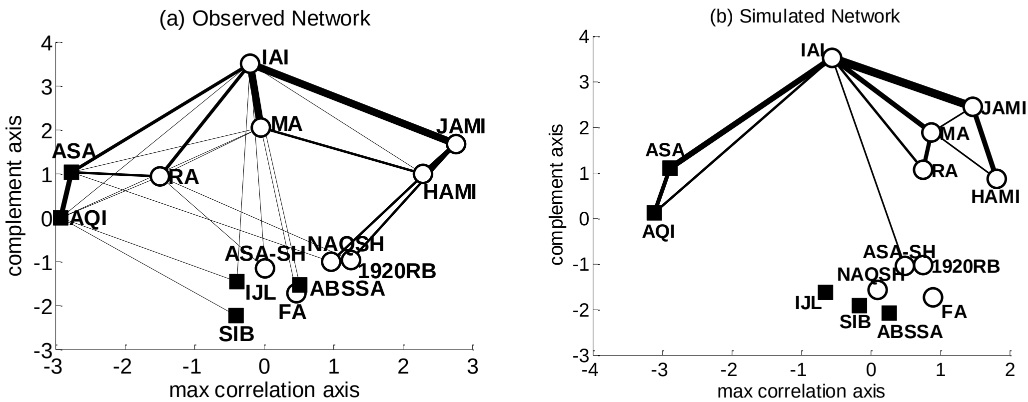

Figure 7. Joint operations networks for Aug07 – Jul09: (a) observed; (b) simulated. Jihadists in black squares, Nationalists in white circles.

Application to Modeling Insurgent Tactical Cooperation

The elements of the above representation can serve as input data for a stochastic model of insurgent tactical cooperation (Gabbay & Thirkill-Mackelprang, 2010). The model variables are the numbers of joint operations between insurgent group pairs, whereas targeting policy and the leadership network ties are taken as fixed parameters. The model assumes that each insurgent group has a number of foot soldier field units that may cooperate with the units of other groups. The model describes the probability that some unit from group *i* will conduct a joint operation with some unit from group *j*. It assumes that the joint operation process depends on the group sizes, the number of prior joint operations between units of *i* and *j*, their tendency to interact with similar other groups, their targeting policies, and the presence or absence of leadership relationships. Figure 8 shows a diagram of the process.

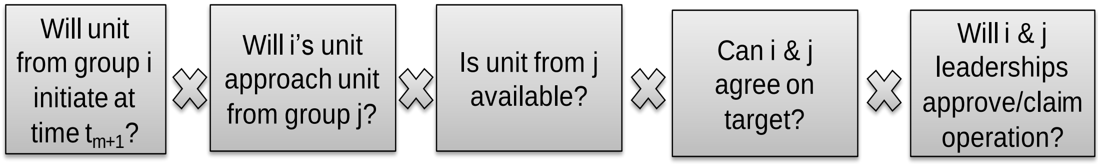

Figure 8. Simulated process for evolution of a joint operations network.

An important feature of this model is its positing of roles for both horizontal interaction in which foot soldier units can, on their own initiative, cooperate with units from other groups and a hierarchical process through which group leaderships can decide whether or not to let such operations go forward and/or claim responsibility for them if they occurred. This horizontal and hierarchical blending manifests the hybrid nature of insurgent groups that mix formal and informal organizational structures.

Figure 7 displays a comparison of the observed and simulated networks for the third time period. The simulated network represents the average over 500 simulation runs. The simulation used the state of the joint operations network in the previous period to represent the initial conditions, but used the targeting policies and leadership network of the third period to better assess the tactical cooperation model itself. Model parameter values (other than the fixed targeting policies and leadership networks) were chosen based on intuitively reasonable estimates and set to be the same for the simulations of the second and third period networks; some variation of parameters was conducted but an extensive search of the parameter space to optimize results was not performed. The visualizations show good agreement in the placement of the insurgent groups; only MA and RA are significantly out of place in the simulation (readers should ignore the links in the simulation plot as they have been thresholded to unclutter the graphic). For forecasting purposes, analysts could employ such a model to predict changes in patterns of insurgent cooperation under different scenarios, such as increased dissension over targeting practices.

7. Future Research Directions

This section identifies some current shortcomings in existing capabilities for data processing in the Forecast capability area and suggests future research directions. By necessity, the discussion is selective and emphasizes areas related to the types of application domains presented in this chapter. It also stresses the need to develop data appropriate for dynamics-based models, advocates more dedicated efforts to collect and process data on elites and their organizations, and suggests the development of multi-level network representations as one avenue for structuring this data.

Data processing is a critical enabling capability in the application of dynamics-based models for forecasting purposes. Better data will naturally lead to better models. While this is true of both statistical and dynamics-based models, the latter ultimately hold more promise. Statistical methods are powerful because they often can be applied to data in a context-independent way, but they have serious limitations when applied to the contexts of particular concern to national security analysts, planners, and policy makers—situations often marked by crisis, conflict, and chaos. For example, regression models can endeavor to forecast future violence levels in a conflict using only a time series of past violence levels without any need to know about the power balance, constraints, and internal dynamics of the warring parties. However, this results in futures that hardly differ from the recent past. Regression models, for instance, would never have anticipated the precipitous drop in violence in Iraq due to the synergistic effects of the Sunni Awakening and the U.S. troop "surge" in 2007, when the Sunnis, squeezed on both sides by the extremist AQI and the newly empowered Shiite majority, decided to side with the United States against AQI (Biddle, Friedman, & Shapiro, 2012). Given that most forecasting of social and political systems occurs within a statistical modeling framework, more focused research is needed on aligning data processing with dynamics-based models.

It is not only the availability of more data that is important but also the intelligent processing of that data into quantities that are (a) theory-driven and (b) operationally relevant. The first criterion refers to the need for compact, low-dimensional quantitative representations that correspond to theoretical constructs from social science. The second refers to the need to include variables and parameters that can change on time scales of operational concern or can be influenced on those time scales using operationally available means. Imposing these criteria will facilitate model development and the practical application of computational models to forecasting. In such applications the models will address the issues most relevant to intelligence analysts and operational planners (thus tying into the Mitigate capability area) and can be used to forecast changes over a relatively small number of dimensions. Operational relevance has not been a strength of most extant instability forecasting work that relies on structural factors such as low economic development, regime type, ethnic fragmentation, mountainous terrain, existence of a territorial dispute, etc. These factors may predispose a country to conflict but indicate little about the situation-specific triggers, timing, and levers of influence that may exacerbate or ameliorate a conflict situation (O'Brien, 2010).

The insurgent network application presented above illustrates a theory-driven quantitative representation of soft, intangible variables for organizations. The definition of insurgent targeting policy – a one-dimensional measure – is motivated by social science research showing the importance of disagreement over targeting practices, such as indiscriminate attacks against civilians, in the relations between insurgent groups. The application of social identity theory to political rhetoric prompted the use of the conflict frame as an ideology variable. The theory's definition in terms of in-groups and out-groups makes the data well suited to serve as input to models of alliance dynamics, as its correspondence to the formation of alliances among Iraqi insurgent groups makes clear. It can also be an input to models of insurgent strategic violence and is in fact correlated with targeting policy (Gabbay & Thirkill-Mackelprang, 2011). The number of parties to the conflict gives the dimension of the conflict frame vector, which can be further reduced by projection onto a master frame space, as described earlier in this chapter. Both targeting policy and conflict frames have operational relevance because they influence and reflect significant changes in the nature of the conflict itself and, hence, evolve on operational time scales. In addition, they can inform analysis and planning of operational means such as information operations or selection of particular groups for targeting or negotiations.

Additional types of data and improved processing techniques focused on elites and their organizations – whether in the government or its opposition, at a national or a local level – would be particularly helpful in developing theory-driven and operationally relevant forecasting methods. The decisions and actions of these individuals and groups directly and immediately shape the trajectories of crises and conflicts. The recent surge of research on popular opinion, driven by the availability of social media and automated tools for sentiment analysis, has overshadowed the significance of elites and their organizations. Although generic sentiment analysis software could be applied to the rhetoric of political leaders and organizations, it would likely not provide a useful data processing tool for forecasting their behavior. Standard sentiment analysis methods, originally developed for contexts such as movie and consumer product reviews, have not proven to work well on the discourse of political elites, whose rhetoric must often be circumspect and cast in neutral language, relying more on nouns and topic emphasis than on adjectives, which carry positive or negative sentiment in ordinary speech (Yu, Kaufmann, & Diermeier, 2008). However, researchers have used other types of text mining algorithms, such as support vector machines, to successfully classify elite political rhetoric (Diermeier, Godbout, Yu, & Kaufmann, 2012).

In support of this call for more research on elite and organizational data, it has been found that the presence of factionalism within a state plays a very significant role in forecasting political instability via statistical methods (Goldstone et al., 2010). This implies that accurate and timely metrics of elite or organizational factional dissension, like the conflict frame-based measure of ideological dissension in Figure 5(b), should improve forecasting of crises such as coups, rebellions, and civil wars. Researchers may find it particularly worthwhile to integrate data on elites with political event data that encode interactions such as violence between political actors. Automated methods for extracting event data from news feeds have been developed for various regions and contexts (Schrodt, 2012). In particular, some researchers have suggested the integration of event data with leadership rhetoric as a way to improve forecasts of the onset of intrastate conflicts such as insurgencies (Tikuisis, Carment, & Samy, 2013).

It is important to stress, however, that forecasting elite and organizational behavior should constitute a goal in its own right and not just an auxiliary input for conflict prediction. This will stimulate the development of such data at a greater temporal resolution, sufficient to track the evolution of these groups and more fully encompass the range of factors that drive their behavior. Elites respond to the preferences of their own organizational constituents, the actions of other organizations and their leaders, and the sentiment of the broader population. Standard social network analysis would focus on constructing ties only within the set of elites or between organizations, but researchers should develop data that enables representation of new multi-level networks, including the ties connecting leaders and the activist or rank-and-file components of their organizations as well as the ties among activists of different organizations.

The use of both leadership tie indicators (joint communiqués) and tactical unit tie indicators (joint operations) represents a step toward better representations of multi-level networks that can more fully account for elite policy and alliance dynamics. For example, it is possible that the standard mechanism of social influence, in which actors with strong ties grow more similar to each other (Friedkin & Johnsen, 2011), appears to be violated at one level but can be explained by ties between levels. Such a dynamic explains the shift of the IAI to a more extreme targeting policy in the August 2007–July 2009 time period (Figure 6), even though its leadership ties in both periods were exclusively with Nationalist groups with more moderate targeting policies, which could have been expected to lead to a moderation of the IAI's targeting policy. This apparent violation of the social influence mechanism can be explained by the fact that the IAI was responding to dissent from its hardline rank-and-file constituents over its recent alliance with JAMI and HAMI (Gabbay & Thirkill-Mackelprang, 2010). The inclusion of the rank-and-file level can therefore account for forces on elites that could produce seemingly anomalous effects when viewed within a single-level network model.

Representations of multi-level networks should also include policies and ideologies. This calls for data that can be used to construct policy and ideological orientations of organizational membership elements. Social media could provide one source of data on rank-and-file opinions, as could analyst judgments. Another approach would involve looking at the rhetoric in media affiliated with elites whose audiences consist of their organizational members or support bases. For example, an analysis of the rhetoric of Afghan government elites found that the statements of the individual leaders themselves were often substantially different, and usually milder, than the rhetoric of their affiliated media, perhaps indicating the more hardline beliefs of their bases of support (Gabbay, 2011). This also highlights the value of incorporating into models data that connect elites with media outlets and media outlets with their audiences.

In conclusion, this chapter has highlighted aspects of the empirical implementation of computational models for forecasting purposes. The discussion has focused on dynamics-based models, which have the greatest potential to address the most operationally challenging and valuable types of forecasting questions but also make the greatest demands upon data processing. Quantitative variables and parameters for low-dimensional models must be coaxed from high-dimensional and often fundamentally qualitative data – a problem compounded by the often contested, chaotic, and covert nature of the environments and adversaries of most concern to national security intelligence analysts, planners, and policy makers. The analyst survey and rhetoric analysis methods described in this chapter represent two different approaches for tackling this problem. In practice, researchers who develop dynamics-based models for forecasting purposes often place a particular model formulation first and then seek to collect and process data appropriate for their model. Although this section has made a few relatively specific suggestions for future research, a broader recommendation would be to pursue a more integrated approach to developing dynamics-based models for forecasting. Such an approach should center on the construction of general quantitative variables appropriate for a particular domain, such as political unrest, insurgencies, or proliferation, which are sufficiently rich to enable the application of dynamics-based models (as opposed to purely statistical models), thereby encouraging a range of approaches that can be more readily evaluated.

References

Alcorn, B., Garces, M., & Hicken, A. (2012). Virthai: A PS-I implemented agent-based model of Thailand as a predictive and analytic tool. *Stanford Journal of East Asian Affairs*, 12, 158 – 170.

Biddle, S, Friedman, J. A., & Shapiro, J. N. (2012). Testing the surge: Why did violence decline in Iraq in 2007? *International Security*, *37*, 7–40.

Bueno de Mesquita, B. (2009). *The predictioneer's game*. New York, NY: Random House.

Diermeier, D, Godbout, J.F., Yu, B., & Kaufmann, S. (2012). Language and ideology in Congress. *British Journal of Political Science*, *42*, 31–55.

Epstein, J. M. (1997). *Nonlinear dynamics, mathematical biology, and social science*. Reading, MA: Addison-Wesley.

Feder, S. A. (2002). Forecasting for policy making in the post-Cold War period. *Annual Review of Political Science*, *5*, 111–125.

Friedkin, N. E., & Johnsen, E. C. (2011). *Social influence network theory: A sociological examination of small group dynamics*. Cambridge, UK: Cambridge University Press.

Gabbay, M. (2007). A dynamical systems model of small group decision making. In R. Avenhaus & I. W. Zartman (Eds.), *Diplomacy games: Formal models and international negotiations*. Berlin, Germany: Springer.

Gabbay, M. (2011). Quantitative analysis of Afghan government actor rhetoric. (Technical report). Seattle, WA: Applied Physics Laboratory, University of Washington.

Gabbay, M. (2013). Modeling decision making outcomes in political elite networks. In K. Glass, R. Colbaugh, P. Ormerod, & J. Tsao (Eds.), *Complex Sciences. LNICST, 126*, 95–110.

Gabbay, M., & Das, A. K. (2014). Majority rule in nonlinear opinion dynamics. In A. Palacios, & P. Longhini (Eds.), *International conference on theory and applications in nonlinear dynamics (ICAND 2012)*. Berlin, Germany: Springer.

Gabbay, M., & Thirkill-Mackelprang, A. (2010, September). *Insurgent operational claims and networks*. Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC.

Gabbay, M., & Thirkill-Mackelprang, A. (2011, September). *A quantitative analysis of insurgent frames, claims, and networks in Iraq*. Paper presented at the Annual Meeting of the American Political Science Association, Seattle, WA.

Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., . . . Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science, 54*, 190–208.

Hafez, M. M. (2003). *Why Muslims rebel*. Boulder, CO: Lynne Rienner Publishers.

Laver, M., & Sergenti. E. (2012). *Party competition: An agent-based model*. Princeton, NJ: Princeton University Press.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, UK: Oxford University Press.

O'Brien, S. P. (2010). Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review, 12*, 87–104.

Schrodt, P. A. (2012). Precedents, progress, and prospects in political event data. *International Interactions, 38*, 546–569.

Tikuisis, P., Carment, D., & Samy, Y. (2013). Prediction of intrastate conflict using state structural factors and events data. *Journal of Conflict Resolution, 57*, 410–44.

Yu, B., Kaufmann, S., & Diermeier, D. (2008, May). Exploring the characteristics of opinion expressions for political opinion classification. In *Proceedings of the 2008 International Conference on Digital Government Research (*pp. 82–91). Montreal, Canada: Digital Government Society of North America.