QUANTITATIVE FRAMEWORK FOR STRATEGIC SPATIAL DECISIONS

Jennifer L. Geffre, Department of Operational Sciences, Air Force Institute of Technology, 2950 Hobson Way, WPAFB, OH 45433, 937-255-3636 (ex4646), jennifer.geffre@afit.edu
Donald E. Brown, Department of Systems and Information Engineering, University of Virginia, 151 Engineer's Way, Charlottesville, VA 22907, 434-982-2074, deb@virginia.edu

ABSTRACT

Supporting the link between risk and strategic decision making is necessary as we continue to strive to deter and disrupt terrorism world-wide. Quantitative risk assessment methods can greatly aid strategic decisions made in an effort to mitigate adversarial threat. This paper outlines a framework for quantitative models in support of strategic decision making. The models predict threat probability and frequency based on the underlying patterns in the multivariate, multilevel, spatial and temporal correlations from previous incidents. It is the combination of results from the quantitative models which aid the high level, long-term planning horizon of strategic decisions. The information gained from applying these models could serve to inform where DIME/PMESII implements might aid in deterring the global terrorism threat.

INTRODUCTION

A key strategy for combating terrorism on a global scale is to "engage in a broad, sustained, and integrated campaign" through the use of military and government instruments including "diplomacy, development, and strategic communications"[9]. Decisions at the operational and tactical level are sufficient for day-to-day operations and short term planning, but provide few insights for meeting long term counter-terrorism goals. Strategic decision makers have Diplomatic, Information, Military and Economic (DIME), and Political, Military, Economic, Social, Infrastructure and Information (PMESII) instruments with which to engage potential targets or potential adversaries. These resources, if employed appropriately, could disrupt the effect of the global terrorism threat in regions most vulnerable.

In understanding which regions are most vulnerable, we must understand what makes them appealing targets. Haimes acknowledges a relationship between the threat to a system and the "states of [that] system" [4]. To parallel that notion, the relationship is assumed to be between the terrorist threat to a region and the conditions within that region (i.e. social, economic, government, military). This connection allows us to formulate predictive models where the conditions become the explanatory variables and the potential for incidents to be the response. In this case, statistical models can identify patterns in previous attack selections and behavior and a threat profile for future activity.

The aim of this paper is to provide an overview of the importance of quantitative analytic methods to strategic decision making and risk analysis and mitigation. The quantitative methods provide powerful insights when they appropriately account for the conditions in a region to include the multivariate, multi-level, spatial and temporal trends of previous incidents. The primary contribution of this work is a robust, comprehensive analytic framework which accounts for multiple sources of correlation in the patterns of incidents. Another important aspect of this framework is its adaptability, both the temporal and spatial structure, in order to apply the framework to various problems related to incident prediction.

The next section discusses the importance of such a framework and the potential to influence strategic decision making, especially surrounding risk. The third section outlines the quantitative analytic approach. The final section highlights some potential uses for this methodology beyond terrorism and military applications.

MOTIVATION

Risk is often defined in terms of probability and consequence [6] or vulnerability and resilience [4]. Risk assessments can be done quantitatively or qualitatively, while risk management incorporates policy decisions and resource allocation to mitigate potential threats. Risk mitigation for threats such as global terrorism should include strategic decisions. These strategic decisions involve long-term plans, which include a high degree of uncertainty; are complex because of the need to consider multiple perspectives and potential mitigation solutions; and ultimately, strive to achieve a goal or set of goals [11].

While it is recognized that intelligent adversaries are adaptive, strategic decisions related to risk mitigation must consider the following questions:

- When and where mitigation efforts are needed?
- Which mitigation implements might be employed?
- How much of each implement is needed?

To address the first question, we must understand where the threat is most prevalent based on previous incidents. We must also consider the conditions of the regions previously targeted, in order to identify similarities to potential targets. We can address the first question in two ways: “what is the probability an attack is likely within a given time window?"; and “what is the expected frequency of attacks in a region within a given time window?".

Fortunately, quantitative methods can aid in determining answers to the above questions. Quantitative models can describe the current environment and highlight unobserved patterns without the bias of decision maker or analyst perception. Quantitative risk studies have begun to included data mining and generalized linear models (GLM), specifically logistic models [8] [12] as a means to evaluate risk. Multivariate models and data mining are well suited to support strategic decision making.

This research focuses on the application of the framework to predicting terrorist incidents worldwide. The incidents were aggregated for each country for a period of one year; hence predictions are made with the same fidelity. While this application focused on large irregularly shaped and irregularly connected regions for one year periods, modifications could be made to accommodate smaller areas (i.e. districts or counties) or for shorter time windows (i.e. quarters, months) depending on the fidelity of the decision to be made.

ANALYTIC FRAMEWORK

The goal of this framework is to aid better strategic decision making, especially in the context of risk mitigation. Analytic models which determine the probability or frequency of potential incidents quantify the vulnerability and allow for specific mitigation opportunities to be investigated. The structure of Logistic and Poisson regression models are modular, where the effects (i.e. multivariate, multi-level, spatial and/or temporal) can be added or removed depending on the underlying pattern of incidents.

The data used for analysis will depend on the subject of the decision being made, the availability of observable characteristics and the fidelity of the analysis required. These different aspects of the problem assume, over a given time window, there is a binary indicator of incident (Y) or a count of the incidents (C). Additionally, each region has a set of individual characteristics or features (X) and features which are pooled based on commonalities among clusters of the regions (Z), a spatial relationship between the regions based on a region adjacency (W) and temporal lags of incidents observed in previous time windows. It should be noted that predictions for a given time window are based on the characteristics or effects of the previous time window (i.e. incident predictions for 2013 would be based on data from 2012; time lagged 1).

Multivariate Effect

The addition of multivariate features in predictive analysis has shown marked improvement in performance [7] [10]. The characteristics or conditions for location i at time t are given as $X_{i,t}$. Multivariate features may include continuous or discrete variables (i.e. demographic, economic, health/medical information) or categorical variables (i.e. indicators for terrorist operations, alliances). The parameter estimates for β determine the importance of each characteristic to the underlying patterns; hence, $\beta X_{i,t-1}$ represents the multivariate effect.

Multi-Level Effect

Studies in social issues and crime [3] [5] have demonstrated how changes to the structure of the model and data to incorporate clusters adds significant benefit in regression models. The multi-level structure of these models account for *fixed* effects (ie. individual location attributes described above as multivariate) and *random* effects (i.e. pooled or clustered attributes). Let G_k (for k = A, B, C, …) represent a grouping or clustering of observations; k could distinguish between different data sources. Within each group, there are subgroups, $S_{[k,l]}$, where k corresponds to the group (or source) and l (for l = 1,2,…) the subgroups associated with k.

The multi-level data for each location i and time t (i.e. data which has been clustered) are represented by $Z_{i,S[k,l],t}$. The multilevel effect includes a random intercept for each subgroup, $\alpha_{S[k,l]}$, and the pooled features, $Z_{i,S[k,l],t-1}$ and the associated features estimates, $\gamma_{S[k,l]}$.This effect is then written as $\alpha_{S[k,l]} + \gamma_{S[k,l]} Z_{i,S[k,l],t-1}$.

Spatial Effect

The spatial effect represents the influence incidents in one region have on its neighbors. Spatial methods which incorporate distance or grids are insufficient to handle large, aggregated irregularly connected locations. The spatial adjacency structure from Anselin [1] accounts for the contiguous relationship between locations. We let W' be the location adjacency matrix, neighboring locations i, j are depicted with a 1, all others are 0. Then, W is the row standardized matrix, where n_i is the sum of each row and $W_{i,j} = W'_{i,j} / n_i$. The spatial effect is then formulated as the product of the spatial significance parameter, ξ, adjacency matrix W and lagged incident vector $Y_{i,t-1}$ (spatial auto-regressive) (i.e. $\xi W Y_{i,t-1}$ for the Logistic model or $\xi W C_{i,t-1}$ for the Poisson model).

Temporal Effect

The temporal effect measures the influence of lagged incidents on the predicted values (i.e. did a previous years' activity continue into the current year?) [2]; time series autoregressive models incorporate easily into the structure of this framework. The temporal effect includes a parameter estimate for each lag, ν, and the lagged incident vectors Y or C^* (where $C_i^* = max\{0.5, C_i\}$). The temporal terms are then written as $\nu_1 Y_{i,t-1}, \nu_2 Y_{i,t-2}, \ldots$ (for the Logistic model) or $\nu_1 \ln\left(C_{i,t-1}^*\right), \nu_2 \ln\left(C_{i,t-2}^*\right), \ldots$(for the Poisson model).

Quantitative Models

The equations which follow show the structure of the effects when combined together. We include a simplifying assumption of homogeneity across the locations. Equations 1 and 2 depict the multivariate, multilevel, spatial and temporal effect structure for the probability and frequency models, respectively. Equations 3a and 3b depict the Logistic regression, which predicts the probability of an incident in region *i* during time *t*. Equations 4a and 4b then show the Poisson model, accounting for the predicted frequency of incidents in region *i* at time *t*.

$$\phi_{Logit} = \beta X_{i,t-1} + \alpha_{S[k,l]} + \gamma_{S[k,l]} Z_{i,S[k,l],t-1} + \xi W Y_{i,t-1} + \nu_1 Y_{i,t-1} + \nu_2 Y_{i,t-2} + \ldots \quad (1)$$

$$\phi_{Poisson} = \beta X_{i,t-1} + \alpha_{S[k,l]} + \gamma_{S[k,l]} Z_{i,S[k,l],t-1} + \xi W C_{i,t-1} + \nu_1 \ln\left(C_{i,t-1}^*\right) + \nu_2 \ln\left(C_{i,t-2}^*\right) + \ldots \quad (2)$$

$$\Pr\left(Y_{i,t} = 1\right) = \pi_{i,t} \quad (3a)$$

$$logit\left(\pi_{i,t}\right) = \phi_{Logit} \quad (3b)$$

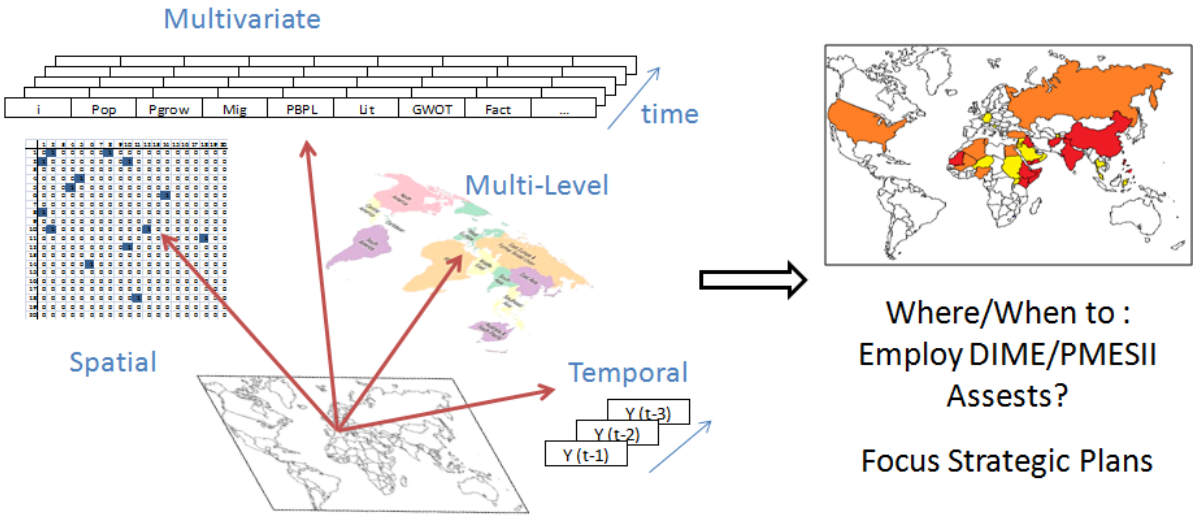
$$\Pr\left(C_{i,t} = c_{i,t}\right) \sim Poisson\left(\lambda_{i,t}\right) \quad (4a)$$

$$\lambda_{i,t} = \phi_{Poisson} \quad (4b)$$

Figure 1 provides a visual representation of the effects incorporated in this methodology. The compilation of all of these effects into a single model aims to include the correlations in the underlying pattern of events and ultimately build a threat profile based on the predictions. These predictions can then inform the strategic decisions regarding planning and resource allocation.

When applied to terrorism incidents worldwide between 2006 and 2011, the preliminary results from this model show nearly a 30% improvement in predictions over a multivariate only model. The misclassification rate (total error) for the Logistic model is less than 10%. Ultimately, by narrowing the focus on areas most vulnerable, we are able to make more informed decisions about the employment of our critical resources.

FIGURE 1



CONCLUSION

The models outlined in this paper provide a quantitative analytic approach for assessing a location's vulnerability to a threat, both in probability and frequency. The incident predictions must include as many potential sources of variation as possible, hence the inclusion of the multivariate, multi-level, spatial and temporal elements of this problem. The prediction of probability informs decision making by identifying potential targets while predicting the frequency aids in understanding the severity of threat.

While the focus of the application of this methodology was attributed to terrorism incidents worldwide, this framework is not limited to this application alone. The framework could influence decisions in the areas of cyber incidents, crime studies, future social unrest, foreclosure trends, sales or marketing impacts. The adaptability makes the framework scalable to any of these examples, given sufficient data is available.

As the US government and military continue to provide diplomatic support, reconstruction and humanitarian efforts, our shrinking budget will require informed decisions for the strategic employment of our precious resources. The patterns uncovered by this framework could ultimately lead to opportunities to preemptively add security measures, improve threat mitigation or attack/incident management policies, and focus aid to struggling countries where they are most vulnerable.

REFERENCES

[1] Anselin, L. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Boston, MA, 1988.

[2] Brockwell, P. & Davis, R. *Introduction to Time Series and Forecasting*, 2nd edition. Springer-Verlag LLC, New York, NY, 2002.

[3] Gelman, A. & Hill, J. *Data Analysis Using Regression and Multileve/Hierarchical Models.* Cambridge University Press, Cambridge, 2007.

[4] Haimes, Y. On the complex definition of risk: A systems based approach. *Risk Analysis*, 2011, 29 (12), 1647-1654.

[5] Huddleston, S. & Brown, D.E. A statistical threat assessment. *IEEE Transaction on Systems, Man and Cynernetics – Part A: Systems and Humans*, 2009, 39 (6), 1307-1315.
[6] Kaplan, S. & Garrick, J. On the quantitative definition of risk. *Risk Analysis*, 1981, 1 (1), 11-27.
[7] Liu, H. & Brown, D.E. Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, 2003, 19 (4), 603-622.
[8] Meng, Q., Weng, J. & Qu, X. A probabilistic quantitative risk assessment model for the long-term work zone crashes. *Accident Analysis & Prevention*, 2010, 42 (6), 1866-1877.
[9] (US) National Security Council. *National Strategy for Combating Terrorism.* Government Publishing Office, Washington, D. C., 2011.
[10] Smith, M. & Brown, D.E. Discrete choice analysis of spatial attack sites. *Information Systems and E-Business Management*, 2007, 5 (3), 255-274.
[11] Wainfan, L. *Multi-perspective Strategic Decision Making: Principles, Methods and Tools*, PhD thesis, Pardee RAND Graduate School, Arlington, VA, 2010.
[12] Whitaker, K.L., Jarvis, M., Beeken, R., Boniface, D. & Wardle, J. Comparing maternal and paternal intergenerational transmission of obesity in a large-based sample. *American Journal of Clinical Nutrition*, 2010, 91 (6), 1560-1567.