Quantitative Assessment in Support of Military Intelligence:
Predictive Analysis and Risk Assessment in Current Security Environment

In Proceedings from the 28th International Symposium on Military Operational Research,
New Place, Hampshire, UK, 28 Aug – 1 Sep 2011

By Dr Peter Dobias[1], Dr Philip T. Eles[1], Dr Jonathan Schroden[2], Dr. James Wanliss[3]

[1]Defence Research and Development Canada, Centre for Operational Research and Analysis
[2]Centre for Naval Analyses, Alexandria, VA
[3]Department of Physics, Presbyterian College, Clinton, SC

Abstract

Present-day conflicts have seen an unprecedented increase in the availability and diversity of quantitative and qualitative information; however, the quality of this information is often less than satisfactory. The available information is frequently one-time reporting, and even for recurrent information methodologies and means of reporting often vary over time. The availability of diverse data sources of varying quality, combined with the complexity of the operational environment makes predictive analysis and risk assessment very difficult.

Possibly the most used and reported information in recent conflicts is reports of various types of enemy and friendly force violent events (i.e. "significant actions"). The internal structure and trends of violence data can provide valuable insights into the dynamics of an analyzed conflict. However, such quantitative assessment needs to be accompanied by a contextual narrative to render it truly informative and relevant.

This paper presents an outline for an analysis of violence data from Afghanistan. It begins with more or less conventional approaches using combinations of various factors and counts of subsets of violence data. It highlights the applicability and limitations of these approaches for predictive analysis and risk assessments. Finally, the paper identifies how fractal-based methodologies can provide insights into internal coherence of the data, identify cross-points, and in general allow an analyst to capture the internal complexity of an assessed environment at multiple scales. These supplementary methodologies can be used to further enhance a situational picture built with conventional analytical methods, and possibly provide enhanced predictive capabilities which might in turn lead to improved risk assessment and management.

Introduction

Present-day conflicts have become an information gold mine. There exist vast amounts of diverse types of data, including reports, numbers, and statistics, as well as other sources of information. However, not all of this information is reliable and useful. In many instances the reports are one-time information. In the case of recurrent reports, the collection methodology, quality control, and even changes in terminology can render analysis of trends fairly difficult. To compound the problem, many users do not consider the quality and reliability of the available data. Often there is an implicit assumption that if the data are available, they must be good. Distinguishing between information sources and assessing the reliability of the available information, compounded with the complexity of the environment with multiple stakeholders of shifting or unknown allegiances and the pressure to present a particular story, makes the work of an analyst a challenge. An additional pitfall threatening the usefulness of analysis is the volume of data available which can lead to a focus on the data for data's sake. Presenting trends and charts can become an end in itself, and the questions "So what?", "What does it mean?", and "What can we do about it?" may go unanswered. And yet, these questions are the *raison-d'être* for analysis.

On the other hand, the availability of diverse information is a great opportunity to gain insight into an evasive enemy who is difficult to track, often uses relatively low-tech capabilities, and has every reason to remain in the shadows. This paper presents several different approaches to analyze the meaning of violence trends and the data structure behind them. Deeper understanding of data structure is a key to more reliable and robust predictive analysis. On the other hand, simple, yet robust approaches to the data can be used to provide a multidimensional picture of the enemy, his preferences and capabilities. This paper uses summary violence data (derived from the significant actions operational reports) as the primary source of information. These data are fairly reliable, well structured, and pliable to interpretation. Some of the discussion will be generic and conceptual, while in some instances it will be possible to dive into the details.

First the paper will discuss different approaches of assessing the enemy's intent and capabilities in a counter-insurgency (COIN) context. Then it will discuss generic properties of data obtained from complex systems, and finally will present an outline of possible approaches to analyzing violence data to obtain deeper understanding of conflict dynamics and to enable predictive analysis in an uncertain environment.

Violence Data in Afghanistan

One of the most complete data sets available is significant actions (SIGACTS) reports, collected and stored in the Combined Information Data Network Exchange (CIDNE) database, mandated for use theatre wide. These event reports are collated and tagged with number of identifiers including event category (e.g. direct fire (DF), improvised explosive device (IED), indirect fire (IDF), or surface to air fire (SAFIRE)), date, time, location, and battle damage assessment (BDA). These data are used for a variety of purposes such as reporting trends (numbers of events changing with time), assessments of the enemy's capabilities, predictive analysis, and finally assessments of internal dynamics of the conflict from a theoretical perspective. This paper will address each of these areas one by one.

While violence data comprise one of the best data sets available, they are far from being perfect. There are delays in reporting due to operational issues, non-uniform reporting practices across different units and different geographical areas, and some reporting involving local forces and civilians is hard to validate. Various estimates of the uncertainties have been made, and in general they are thought to be up to 10% of the total number of violent actions in a given time period

([1],[2]). This needs to be kept in mind when analyzing trends. On the bright side, the uncertainty is generally in under-reporting, so the temporal trends, in particular long-term trends, are fairly robust.

The most common means of presenting Afghanistan violence data is to show weekly or monthly summary data over time. Figure 1 shows an example of monthly overall violence numbers. In some cases the data may be broken down by category. While common, this way of presenting data is grossly inadequate, because it does not provide any insight into the context of the violence. Is the increase or decrease because of external factors (weather, climate, harvest, or holidays), or internal factors such as a shift in insurgent or counter-insurgent capabilities or intent? Is it because of a change in one side’s operational tempo? Is there a geographical context to the violence? How much of the violence is due to crime, or due to local grievances that have nothing to do with the larger insurgency? From these unanswered questions it is obvious that this way of looking at the data is inadequate.

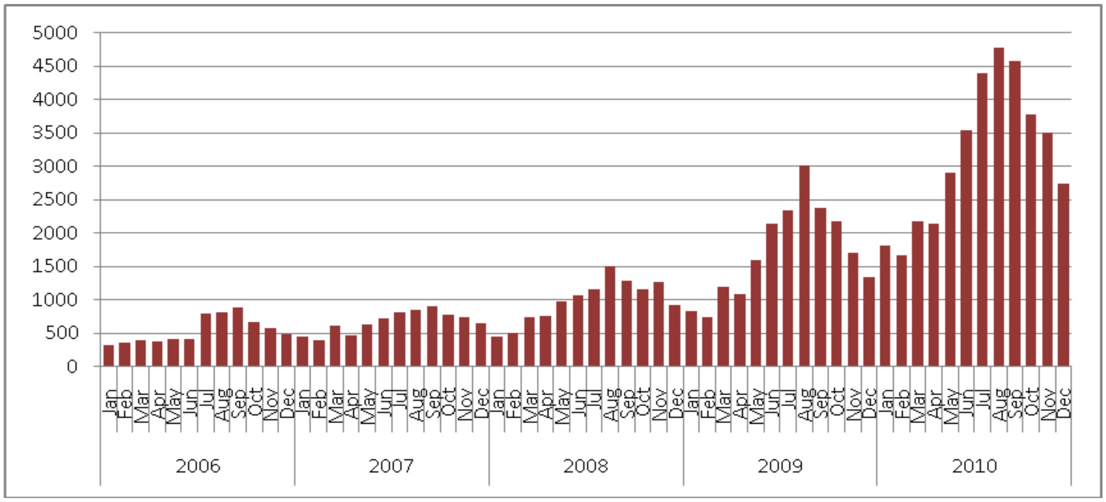

Figure 1. Monthly violence data from Afghanistan.

There are two noticeable trends in the data in Figure 1. One is almost periodic, an annual cycle of increase and decrease, while the other is some kind of year-to-year increase. The annual cycle is connected with the climate. The winter months render passage of the mountainous areas at the border between Pakistan and Afghanistan as well as movement and operations in eastern Afghanistan difficult; therefore violence tends to wind down slightly during the winter months. This is compounded by the fact that insurgents have their own annual recovery and resupply cycle that coincides with the annual climatic cycle. The annual increase is more difficult to explain; one of the driving factors might be an increase in the coalition forces’ footprint over time. To separate these two major trends, a methodology called *seasonal decomposition* [3] can be used.

To briefly summarize the seasonal decomposition technique as it has been used in violence analysis, the idea revolves around factoring a time series of data X_t into two parts: a seasonal component (S_t) and a trend component (T_t). A multiplicative model was used for the analysis [4]:

$$X_t = T_t \times S_t$$

If the cycle can be known or estimated *a priori*, the individual components can be calculated as follows:

- First, a moving average χ_t is calculated for the series X_t, with the width of the window equal to one season (period). This removes all intra seasonal variability. For the Afghanistan data, the assumed season/period was 12 months.
- Next, the ratio of the observed X_t and the smoothed series χ_t will isolate the raw seasonal component. The average seasonal component S_t is then calculated as the normalized medial average for each point in the season.
- Finally, the trend component T_t is obtained by dividing the original series X_t by the seasonal component S_t.

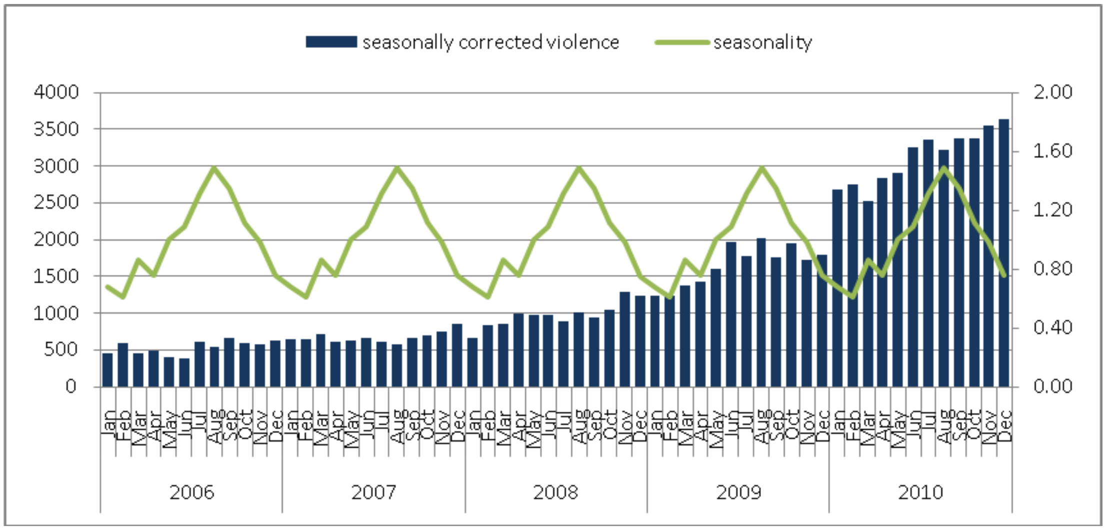

Figure 2. Seasonally corrected violence in Afghanistan.

Figure 2 shows the results of seasonal decomposition on the Afghanistan violence data in Figure 1. The annual increases in violence (blue boxes), after the seasonal variations (green line, right axis) are removed. The increase appears to be exponential between 2006 and 2010 ($\propto \exp(0.09 \times \mathrm{month})$, with R^2=0.94)). A slight slowing down of the increase appeared toward the end of 2010. This long-term seasonally-corrected trend can be correlated with other longer-term changes, and subsequently used for predictive analysis. For example, the increase is well correlated with the increase in coalition troop numbers (R^2~0.90) during the same timeframe. This does not imply causality, but it implies that troop numbers can be used as a proxy for an estimate of violence. One of the possible explanations for this is that the system behaves like an unstable system with a positive linear feedback[1]. The feedback can be explained by a simple loop: increased violence leads to coalition forces putting in more resources, leading to a larger footprint and thus more violence, and so on. There were other explanations proposed, including the spread of insurgency via splinter groups, or via revenge-based

---

[1] A system with a linear positive feedback is akin to a harmonic oscillator, but the force is positive rather than negative ($\ddot{x} = kx$). The solution for such a system is an exponentially increasing displacement (ballooning instability: $x \propto x_0 \exp(kx)$).

increased insurgent footprints [5]. These mechanisms are not mutually exclusive, and likely each of them has some contribution to the trend.

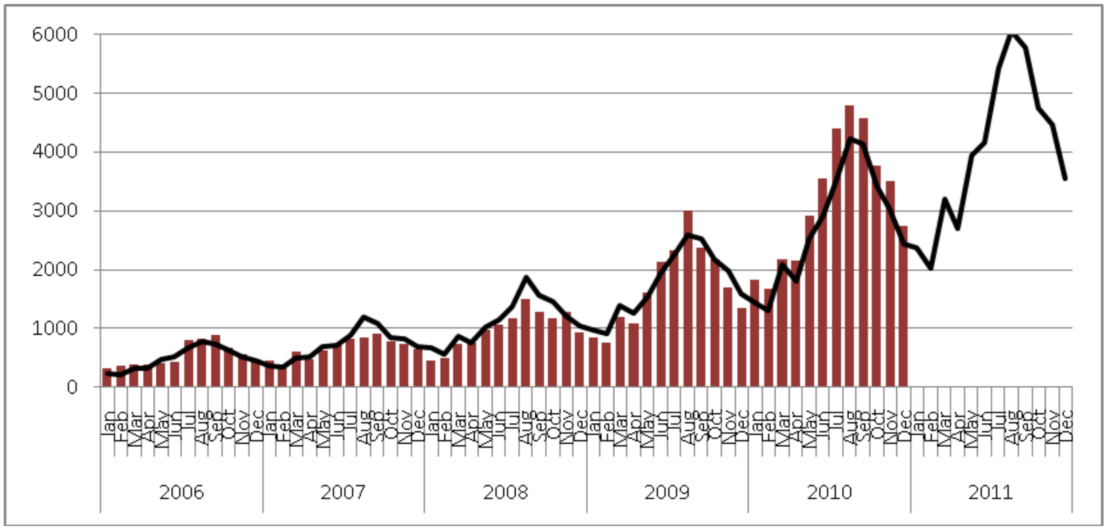

Figure 3. A notional estimate (black line) of the violence levels for 2011 based on security force numbers and the historical relationship between force levels and violence.

Figure 3 shows a notional estimate of future violence levels based on the historical relationship between security force numbers and violence, assuming a continual exponential rise which presumes that nothing has changed in the conflict dynamics. This is overly simplistic assumption, and a realistic estimate would have to include additional considerations and would need to identify any shifts in the relationships that may have occurred.

Insurgent Capabilities and Intent

Of particular interest to the intelligence community is understanding the insurgency, especially insurgent intent and capabilities. For example, is an increase in violence in a particular region due to a change of insurgents' focus, or because they wanted the counter-insurgent forces to divert capabilities from their real focus area? What impact do coalition operations have on the insurgents' capabilities and resources? Are insurgents losing their grip on the population? Analyzing different violence categories and the relationships between them and their geographical distribution can provide such insights [6]. This information must be used in the context of other available information to corroborate or disprove the conclusions made directly from violence data.

Among methodologies that can be used are correlations between different subcategories (e.g. total violence and effective attacks), analysis of covariance, trends and shifts in outliers, etc. This information can provide useful insights that can be corroborated with other sources to paint a picture of the conflict from multiple perspectives. Inclusion of various perspectives of the violence is important, since both the data and an analyst's situational awareness are limited, and no single measure provides an adequate picture.

Geographical Shifts in Data

Geographical distribution of violence (Figure 4) can provide additional interesting insights into the conflict in Afghanistan. It is heavily driven by the topography of the region (as is everything in Afghanistan, including population distribution). Some of the key observations are:

- The violence is concentrated in the southern and eastern regions
- The distribution is largely focused in the vicinity of the Ring Road (major road connected various parts of Afghanistan)
- The violence in eastern Afghanistan is more widely distributed than in the south.

Temporal changes in the geographical distribution can provide further insights into the spread of the insurgency, and insurgent focus areas. In particular, if the geo-temporal trends are correlated with other activity such as major operations, seasonal events such as harvests, etc., they can provide good insight into the interaction between insurgents, the population, and security forces. Additional information can be gleaned from analysis of geographical distribution of particular types of violence (e.g. direct fire and/or improvised explosive devices).

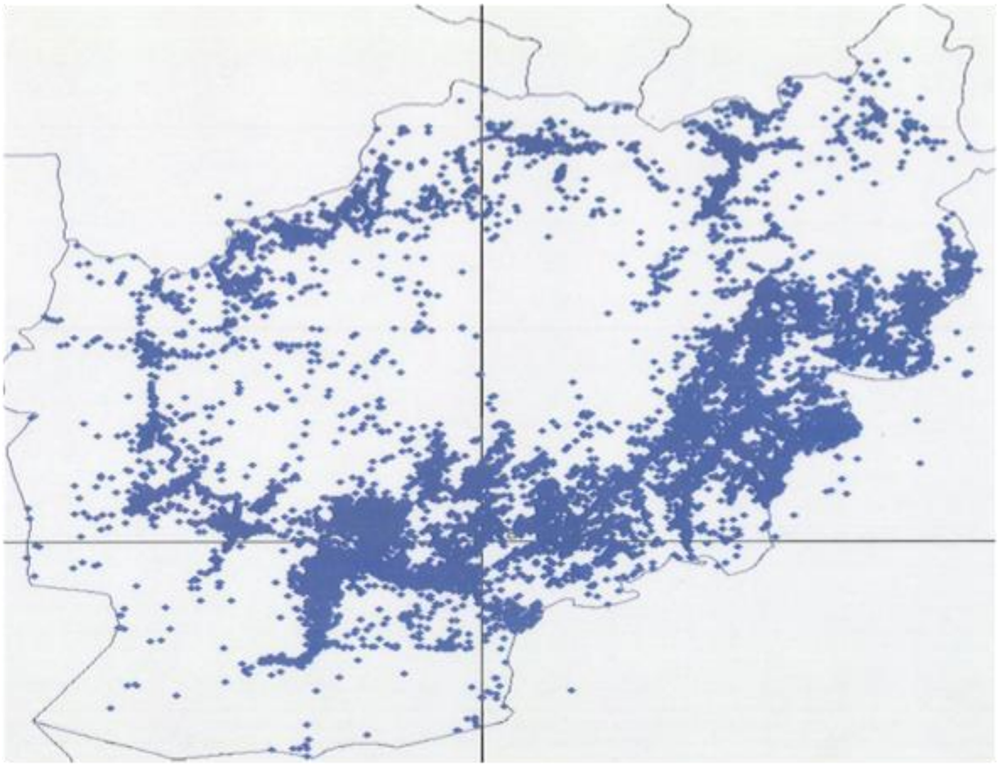

Figure 4. Geographical distribution of violence in Afghanistan over the past 3 years.

Among the methodologies applicable to geo-analysis of violence in Afghanistan are density and density differential plots. Often the question arises whether it is better to map points or densities. Points can show exact location of events, but the densities provide better insight into the actual number of events. For example, the point map (Figure 4) does not allow for a comparison of violence numbers in Helmand and Kandahar, because the high number of events results in a lot of overlapping points.

A density differential enables a comparison of relative numbers of violent events in a particular area. It needs to be used in conjunction with a point or “real density” map, since a large change does not imply large activity and vice versa. Relative changes in the densities of various types of attacks may reveal useful information about enemy focus areas and so prevent focusing resources on decoy areas.

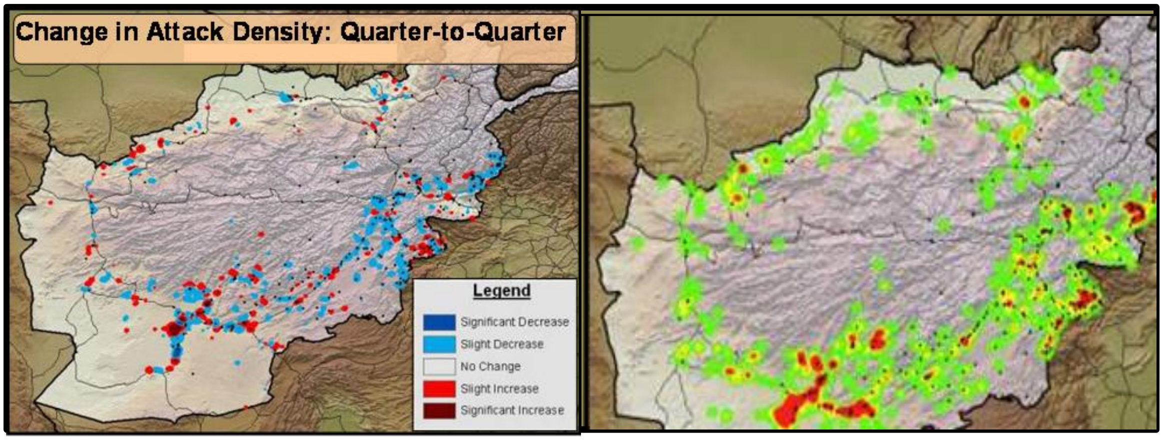

Figure 5. Example of a density differential map with the actual density in the left-hand corner.

Complex Adaptive Systems, Indices, and Fractal Distributions

It has been shown before that the casualty numbers in Afghanistan obeyed fractal (power law) distributions [7] and were intermittent [8], behaving in a way consistent with a fractal point process [9]. Similar analyses can be made in terms of violence numbers. However, this is more problematic, due to the two main trends outlined above (i.e. seasonal and year-to-year fluctuations). These trends need to be removed from the data to isolate the behavior of deviations from “normal” behavior. If the trend is not removed the scaling would become dependent on the time window (i.e. in the case of Afghan violence the large numbers of events per day are much more likely to occur closer to the present than they were few years ago (Figure 6)).

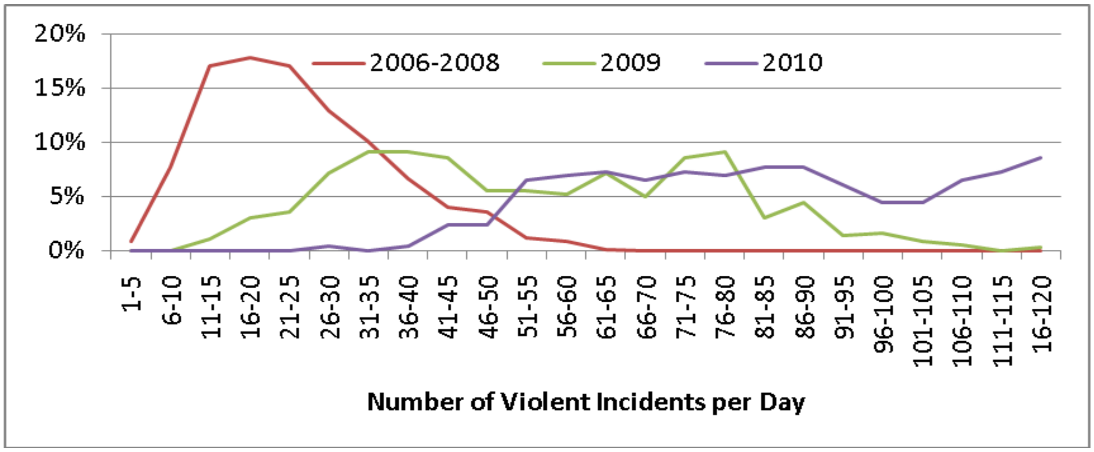

Figure 6. Distribution of violent events in Afghanistan 2006-2010.

One way of doing this analysis is to look at the scaling of the square deviation from the seasonally decomposed trend (Figure 2). A 4th order polynomial was identified as the best fit (R^2 ~ 0.70):

$$N = -3.4 \times 10^{-11}x^4 + 1.575 \times 10^{-7}x^3 - 2.028 \times 10^{-4}x^2 + 9.899 \times 10^{-2}x$$

The polynomial rather than exponential fit was selected in order to capture the change in the trend observed since late fall 2010.

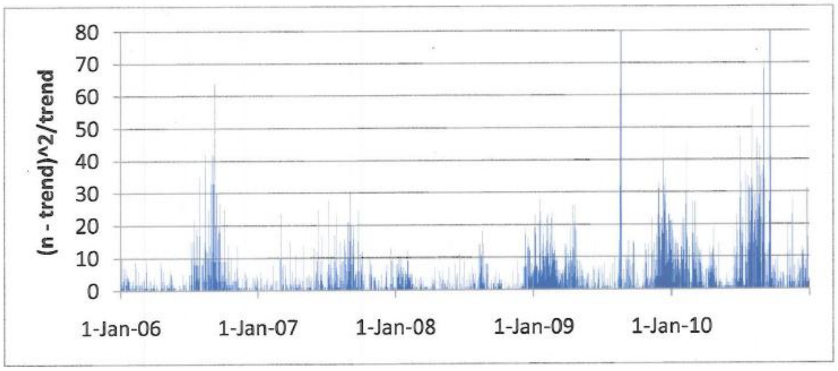

Figure 7. Temporal dependence of square deviations.

Once the relative square deviation (Figure 7) from the trend was calculated, the frequency-size distribution was calculated (Figure 8). The distribution obeys a power law (as was expected on the basis of earlier studies [10]) with the exception of very small deviations (they were less frequent than would be expected from the power law). The scaling coefficient was γ = 2.42, which is consistent with earlier studies [11]. If the distribution is assumed to be bi-fractal (different from lower and higher deviations), the fit can be further improved, with the scaling coefficient γ ~ 1.8 for smaller deviations and γ ~ 2.8 for larger deviations. Since the deviations are not distributed uniformly (Figure 7), the different scaling might be related to the seasonality of the violence.

The fractal scaling of the incident frequency has potential significant implications for the predictive analysis and modeling of this type of conflicts. The key implication for the predictive analysis is that changes in the system are not random, but are result of a long-term “memory” (i.e. dependence of the system on earlier variations). Therefore any stochastic (e.g. Monte-Carlo) methods that would rely on the normal distributions of the variations would be inadequate. In addition, the second key implication for both predictive analysis and modeling is a strong dependence on the initial conditions that may render forecasting unreliable. On the other hand, if there are attractors in an analyzed system, it might be possible to estimate probabilities of alternative futures without detailed knowledge of all the relationships in the system.

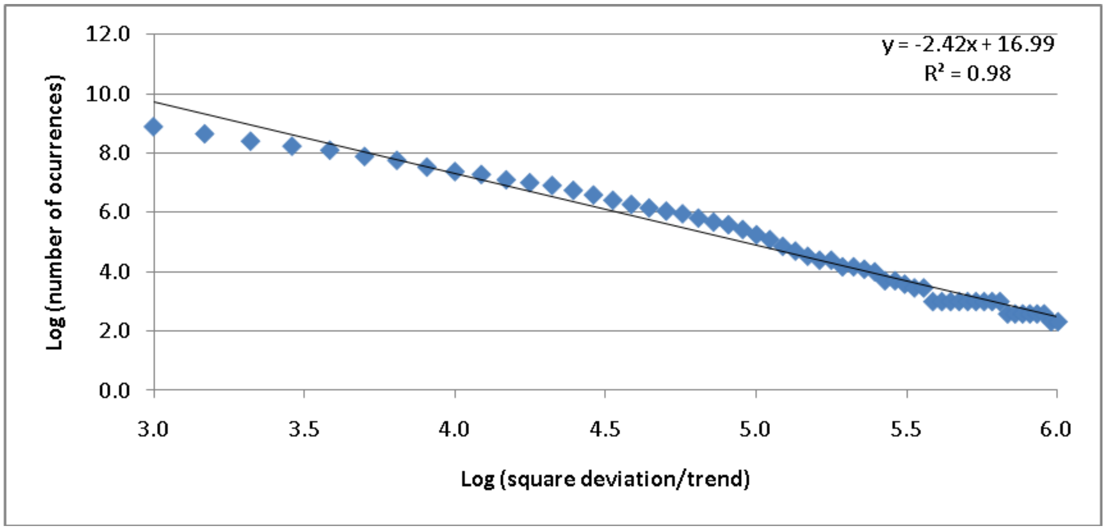

Figure 8. Log-Log plot of the frequency-size distribution for the square deviation from the polynomial fit.

Currently the focus of the research by the authors is on revisiting the analysis using more complete data (e.g. references [7] and [8] used only 18 months of data available at the time. In addition, relationships between fractal properties of incidents and casualties are being explored. The objective of this line of inquiry is to enable effective modeling of irregular warfare using agent-based models.

Predictive Analysis

Of particular interest for the consumers of the analysis is forecasting what is likely to be the violence level in the future. This is important to manage expectations, prepare sufficient resources to mitigate losses, have sufficient medical and transportation capacity and appropriate distribution of medical facilities, etc.

Due to a large number of unknowns and limited understanding of the dynamics of the conflict, only limited forecasting is possible. As was already mentioned above, one of the good indicators for future violence has been friendly force numbers. Since there has been a shift in that trend in 2010, it is currently unknown if this remains a reasonable indicator. Because many variables have changed in the course of 2010, it is virtually impossible to forecast violence levels on the basis of the historical trends alone. However, it was demonstrated previously that the violence data between 2006 and

2009 exhibit long-term memory [10]. In other words, the violence level at any given time is correlated with the violence in proximal time. This property can be used to develop methodologies to predict future behavior.

Recent decades have witnessed the development of new tools that allow for the analysis of data that feature sudden transitions and extreme events, similar in nature to the violence data from Afghanistan. These advances have been made possible by a combination of approaches based on dynamical systems theory and statistical physics. Fractal and multi-fractal approaches have been quite successful in extracting salient features of various physical processes. These approaches have been criticized because they are often perceived to be anathema to the traditional way of approaching problems. Instead of reductionism, the approach of complex statistical physics insists on a holistic methodology, one that considers the statistical interactions as well as the parts. The necessity of such statistical analysis is rooted in the fact that, in general, nonlinear systems with multiple spatially distributed sources of instability cannot be completely characterized in deterministic terms. A significant portion of information is contained in multi-scale correlations of non-Gaussian random variables. These depend on the system dynamics, which can be extracted by appropriate statistical-physical methods of analysis. Two such methods are described below.

Waiting Times

One of the recent methods in use in the study of extreme natural data is the time interval between successive events above (or below) some threshold, sometimes referred to as re-occurrence times, waiting times, or return intervals. By studying the statistics of the waiting times for increasing threshold heights, one aims to find out the laws governing the occurrence of extreme events.

Since the statistics of waiting times between extreme events in real systems is quite poor one usually tries to extract information from events with smaller magnitudes that occur quite often and thus have enough statistics. The major issue is to find out some general ‘scaling’ relations between the return intervals at low and high thresholds, which then allow one to extrapolate the results to very large, extreme thresholds. It is usually assumed that extreme events, such as a major offensive in military terms, are uncorrelated due to the large time span in between events. Under this assumption the probability density function of the waiting times is a simple exponential, representing a Poisson process,

$$P(\tau) = A\exp\left(\frac{-\tau}{<\tau>}\right)$$

where $<\tau>$ is the mean waiting time, and the individual waiting times are uncorrelated. Accordingly, the extreme events are randomly distributed. This result is, however, not always in agreement with the observations; extreme events in natural systems often occur in clusters ([12],[13]). Breaking of symmetry can be interpreted as a regime change in the system dynamics. Again, in the military context, it means investigating whether breaking of symmetry suggests that a change in combat dynamics is imminent.

Forecasting via Multi-Fractal Symbolic Dynamics

Because of the fractal nature (as witnessed by the power-law scaling and persistence) of combat dynamics it may be possible to leverage data statistics to make probabilistic hazard assessments for combat dynamics, as will be explained below. These are primarily based on the association of small disturbances with future large outbreaks of hostilities. Nonlinear and multi-fractal models typically feature a sensitive dependence on initial conditions which means that a direct quantitative comparison between real and simulated data is complicated. For this reason one usually turns to measure representations of data. Below is an outline of a technique to compute such a measure representation, based on the method of *Yu et al*. [14] and *Wanliss et al*. [15].

The idea is to convert the signal into a symbolic string sequence using a thresholding algorithm. This is essentially a representation of the time series in the form of a probability measure of *k*-strings. The traditional way of calculating a probability density function is to sum all occurrences of a particular value in a time series. The measure representation essentially generalizes this process by considering not just one value, but a string of values of length *k*. The probability that this string occurs in the entire time series is then computed in a method analogous to that of the ordinary density function. For violence data, one possible approach is to consider the series to consist of data representing two types of behaviors, namely, quiet (Q) or active (A). The definitions may be defined in terms of no casualties (Q), or casualties (A), or in other terms, such as violence above a certain threshold of tolerance, etc.

Once the threshold has been defined, each data point in the time series is classified according to one of these two possible behaviors (letters). A new time series can then be computed from the original. Strings of length *k*, of consecutive Q or A values, can be computed. Any string made up of *k* letters from the set {Q,A} is called a *k*-string. For a given integer number *k* there are a total of 2^k different *k* strings. In other words, in order to count the number of each kind of *k* strings in the D_{st} time series, 2^k counters are needed. The interval *[0,1]* is divided into 2^k disjoint subintervals, and each subinterval is used to represent a counter. Let $s = s_1,..., s_k, s_i \in \{Q, A\}, i = 1,..., k,$ be a substring with length *k*; one then defines:

$$x_i(s) = \sum_{i=1}^{k} \frac{x_i}{2^i}, \text{ where } x_i = \begin{cases} 0 & if\ s_i = Q \\ 1 & if\ s_i = A \end{cases} \text{ and } x_r(s) = x_l(s) + \frac{1}{2^k}.$$

Next, the subinterval $[x_l(s), x_r(s))$ is considered to represent the substring *s*. Let *N(s)* be the number of times substring *s* appears in the D_{st} time series. Then

$$F(s) = N(s) / L^k$$

is the frequency of substring *s*. It follows that $\sum_{\{s\}} F(s) = 1.$ It is now possible to view *F(s)* as a function of *x* and define a measure μ_k on *[0,1]* by

$\mu_k(x) = Y_k(x)dx,$ where $Y_k(x) = 2^k F_k(s),\ x \in [x_l(s), x_r(s)).$

In this case $\mu_k([0,1)) = 1$ and $\mu_k([x_l(s), x_r(s))) = F_k(s).$ $\mu_k(x)$ is the measure representation of the data set. Once the measure is completed, probabilistic forecasts may be attempted, assuming that the nature of the conflict dynamics does not change.

The ability to predict violence levels with any reasonable accuracy for at least a short time ahead (similar to weather forecast) would enable more efficient resource allocation and expectation management. More work remains to be done in order to determine feasibility of the outlined methodology. Of particular concern is the size of the available dataset. The above methodology was developed and tested for systems that yield millions of available data points, and thus the limited SIGACTS data set might not allow for reliable forecasting.

The planned research along this line of inquiry is looking at re-analyzing persistency within SIGACTs using longer data set. Then the data will be used to test the model hypothesis and the feasibility of the analysis for the SIGACT data.

Summary and Conclusions

While not perfect, violence data are of interest to the intelligence community. They can provide interesting and potentially important insights into the nature of the insurgency that can corroborate other information sources, and sometimes can direct intelligence analysis. In addition, violence data can be used to deepen understanding of the nature of a conflict, to plan resources and to prevent needless diversion of resources to areas of secondary importance.

In some instances, like in Afghanistan, seasonality can play an important role. In such case statistical methods such as seasonal decomposition provide means of eliminating seasonality from the picture to obtain a long-term trend in the violence. Geo-analysis of the data can provide important insights into the special distribution of a conflict, and in particular into the focus areas for the enemy. Distinguishing real focus areas from attempts to divert counter-insurgent resources is vital if there are significant resource constraints in place, as is often the case.

Understanding the scaling nature of the violence data has the potential to enable at least limited predictive analysis that can be useful for the management of resources in the counter-insurgency. More work is required on the development of multi-fractal predictive techniques, and in particular their application to warfare models.

References

[1] IJC CIDNE Collation Cell, Personal Communications, June 2010

[2] P. Dobias, Estimating Uncertainties in the SIGACT Numbers, Decision Paper, USCENTCOM Afg-Pak Center, 8 Oct 2010

[3] StatSoft, Inc. Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/ (http://www.statsoft.com/textbook/time-series-analysis/#census), 2011.

[4] J. Schroden, "Real" Violence Trends in Afghanistan: Taking Seasonality out of the Picture, US CENTCOM, Feb 2011

[5] J. Schroden, Analysis of Violence Contagion in Afghanistan, US CENTCOM, March 2011

[6] Assessing the Insurgency in Afghanistan: A Model and Indicators, US CENTCOM AFG-PAK CENTER, March 2011

[7] P. Dobias, Self-Organized Criticality in Asymmetric Warfare, Fractals, Vol 17, No 1, pg 91-97, 2009

[8] P. Dobias, K. Sprague, Intermittency of Casualties in Asymmetric Warfare, Journal of Battlefield Technology, Vol. 12, No. 1, 19-25, 2009

[9] D.R. Bickel, "Estimating the Intermittency of Point Processes with Applications to Human Activity and Viral DNA", *Physica A*, Vol. 265, pp. 634-648, 1999

[10] J. Wanliss, P. Dobias, K. Sprague, Persistence in Significant Activities in Afghanistan, Journal of Battlefield Technology, Vol. 13, No. 3, 25-29, 2010

[11] J.C. Bohoroquez, S. Gourley, A.R. Dixon, M. Spagat, N.F. Johnson, "Common ecology quantifies human insurgency", Nature, Vol. 462, No. 17, 2009, pp 911-914, DOI:10.1038/nature08631

[12] A. Bunde, Long-term persistence in nature: On the clustering of extreme events, in: 'Aha Hulikoa'a' Workshop on Extreme Events, Proceedings of the Hawaiian Winter Workshop in Manoa, Hawaii, pp. 19 – 25., 2007

[13] A. Garber, H. Kantz, Finite-size effects on the statistics of extreme events in the BTW model, European Physical Journal B 67, 437-443.2009, 2009

[14] Z.G. Yu, V. Anh, K. S. Lau, Measure representation and multifractal analysis of complete genome, Phys. Rev. E, 64, 31903, 2001

[15] J.A. Wanliss, V. V. Anh, Z. G. Yu, and S. Watson, Multifractal modelling of magnetic storms via symbolic dynamics, J. Geoph. Res., Vol. 110, A08214, doi:10.1029/2004JA010996, 2005