Assessing the influence of agricultural production on the Taliban Insurgency: A spatial and temporal lag approach

Kevin Stofan

GEOG 596A Peer Review
Spring 2011

Agenda

- Background and Purpose
- Structured Analytic Techniques
- Study Area and Data
- Exploratory Spatial Data Analysis (ESDA)
- Spatial Regression Modeling

Figure 1. Arghandab River valley, Afghanistan and pomegranate garden.

Background

- Observation of increased recruitment in low agricultural production areas.

- Improvement on the reporting of raw significant activities as a measure of relative violence (use of rates).

- Accounting for the presence of security forces as an explanatory variable for insurgent attacks.

Figure 2. Arghandab River valley showing village and typical adjacent agriculture.

Spatially and Temporally Lagged Violence

- Taliban less likely to conduct attacks within their home villages.
- Fighting age males without work due to agricultural shortfalls susceptible to local, short-term recruitment
- Fighters and low/mid-level commanders associated with productive areas display larger temporal lags due to planning, travel, and organization.

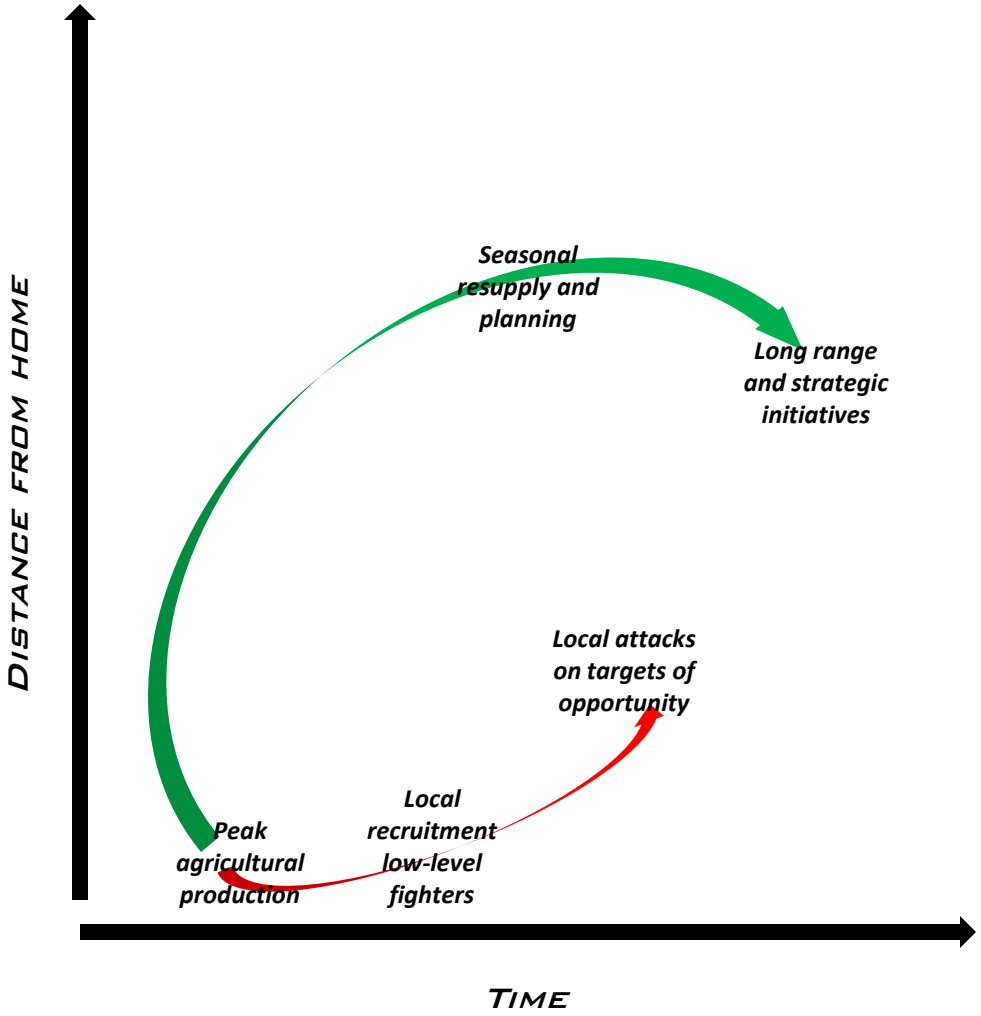

Figure 3. Conceptualization of Taliban recruitment, planning, and attack cycles.

Purpose and Methodology

- The purpose of this study is to determine if low agricultural production is related to increases in violent events at the village level.

- Uses structured analytic techniques, ESDA, and spatial regression within an SGAM framework.

- SGAM provides an analytic framework for complex geospatial problems that promotes critical thinking and limits bias in intelligence research.

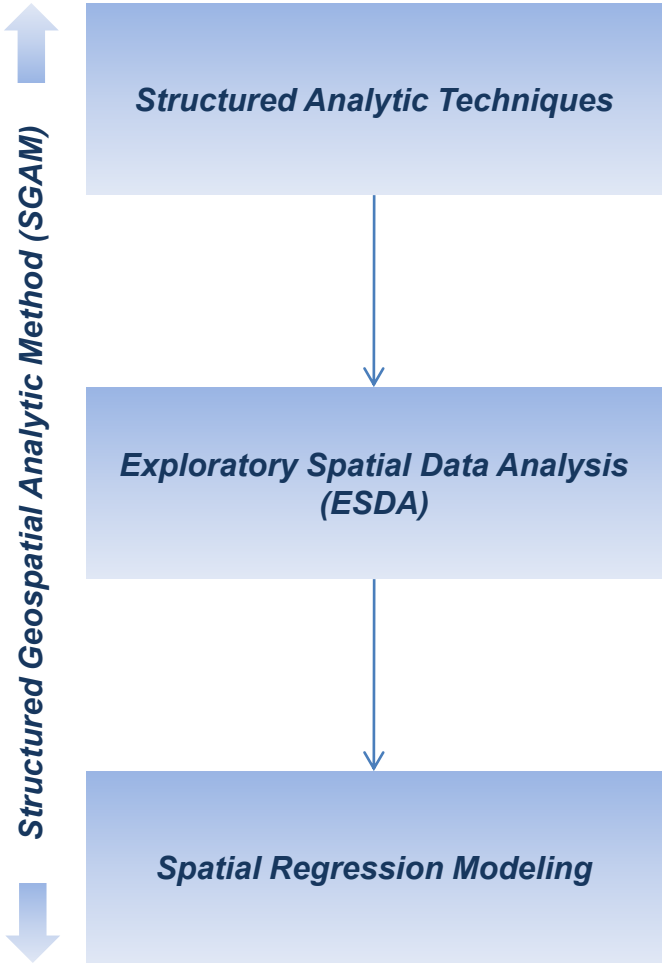

Figure 4. Proposed analytic framework.

Structured Brainstorming

- Used to increase the pool of knowledge used to identify assumptions, explanatory variables, and alternative hypotheses.
- Two week session consisting of a divergent and convergent phase.
- Diverse group of participants including scholars, military, intelligence analysts, and native Pashtuns.

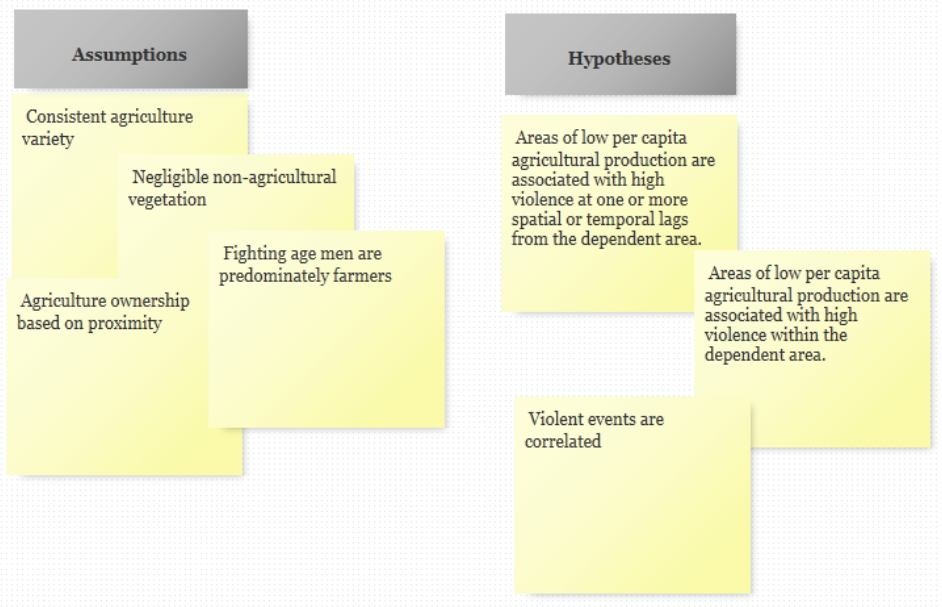

Figure 5. Example of structured brainstorming from stixy.com

Quadrant Crunching

- Creates a robust list of stories by developing contrary (opposite) assumptions and constructs every possible combination.

- Used where little data is available and the chance of surprise is high.

- Uses inputs from the structured brainstorming to create a larger set of assumptions, variables, and storylines.

Key Assumptions	Contrary Assumptions
Population density does not vary within the district.	Population density does vary within the district.
Proportion of structure area to irrigated crop area is equal within the district.	Proportion of structure area to irrigated crops varies within the district.
Features extracted from Quickbird are accurate representations.	Features extracted from Quickbird imagery are not accurate representations.
Settlement centroid and associated Thiessen polygon is the best aggregation of populations.	Settlement centriods and polygons are not a logical aggregation of populations.
Minority Durrani population has the same economy/lifestyle as Ghilzai majority.	Minority Durrani population has different economy/lifestyle than Ghilzai majority.
Varied Population Density / Structure:Crops Varied	
High birth rate / Herder population	High birth rate / Xerophytic crops
Low birth rate / Herder population	Low birth rate / Xerophytic crops
Varied Population Density / Feature Extraction Errors	
High birth rate / Structures uninhabited	High birth rate / Ground interference
Low birth rate / Structures uninhabited	Low birth rate / Ground interference
Varied Population Density / Thiessen Polygon Incorrect Aggregation	
High birth rate / Multiple tribe settlement	High birth rate / Within settlement variance
Low birth rate / Multiple tribe settlement	Low birth rate / Within settlement variance
Varied Population Density / Durrani:Ghilzai Economy Difference	
High birth rate / Durrani sedentary	High birth rate / Rurality variance
Low birth rate / Durrani sedentary	Low birth rate / Rurality variance
Structure:Crop Varied / Feature Extraction Errors	
Herder population / Structures uninhabited	Herder population / Ground interference
Xerophytic crops / Structures uninhabited	Xerophytic crops / Ground interference

Figure 6. Example of quadrant crunching output.

Analysis of Competing Hypotheses (ACH)

- Used in situations where alternative explanations of a process are likely.
- Avoids satisfying an initial or leading hypothesis by refuting hypotheses.
- Uses the results from structured brainstorming and quadrant crunching to develop alternate hypotheses.
- Hypothesis with the least evidence against it is chosen as the most plausible.

		Credibility	Relevance	H: 1	H: 2	H: 3	H: 4
				Population is randomly dispersed in rural Pashtun Afghanistan and not significantly correlated with any geospatially derived features.	Settlement area in rural Pashtun Afghanistan is significantly positively correlated with population.	Settlement vegetation area in rural Pashtun Afghanistan is significantly positively correlated with population.	Distance from urban areas of rural Pashtun settlements is significantly negatively correlated with population.
	Weighted Inconsistency Score ➪			-4.0	-1.707	-0.707	-10.0
	Enter Evidence						
E12	Settlements are self sufficent and distance to a district or provincial center does not provide much of an advantage to a settlement.	MEDIUM	MEDIUM	C	N	C	II
E11	Limiting factor of certain settlement's amount of arable land is terrain not manpower.	LOW	HIGH	I	C	C	I
E10	Time to construct dwellings is labor intensive and typically only occurs if needed.	HIGH	HIGH	N	C C	N	N
E9	Some appearance of quality of life indicator variation did exist.	MEDIUM	LOW	N	I	I	C
E8	Some dwellings in settlements may be abandoned or grain storage areas.	HIGH	LOW	N	I	C	N
E7	Larger settlements (by area) have a higher likelyhood of having higher order functions i.e. bazaars or clinics.	HIGH	HIGH	I	C	N	I
E6	Settlement placement is dictated by physical terrain.	MEDIUM	MEDIUM	I	N	N	I
E5	Little to no variation exists in the structure and design of settlements between settlements.	HIGH	HIGH	N	C	N	N
E4	No discernable pattern of varied population with tribal affiliation exists.	LOW	MEDIUM	C	C	N	N
E3	Settlements do not show any patterned variance in size with distance from district or provincial center.	MEDIUM	MEDIUM	C	C	N	II
E2	Larger settlements typically have more agricultural land area.	HIGH	HIGH	N	C	C C	N
E1	Multiple observers have noticed increased numbers of inhabitants with increased size of a settlement.	HIGH	HIGH	N	C C	N	I

Figure 7. Example of analysis of competing hypotheses (ACH) output.

Study Area

- Located in north central Kandahar province.
- Dominated by the Arghandab river valley flowing from northeast to southwest.
- Directly adjacent (NW) to Kandahar city.
- Relatively strong economy based on agriculture (pomegranate, grapes, spices).

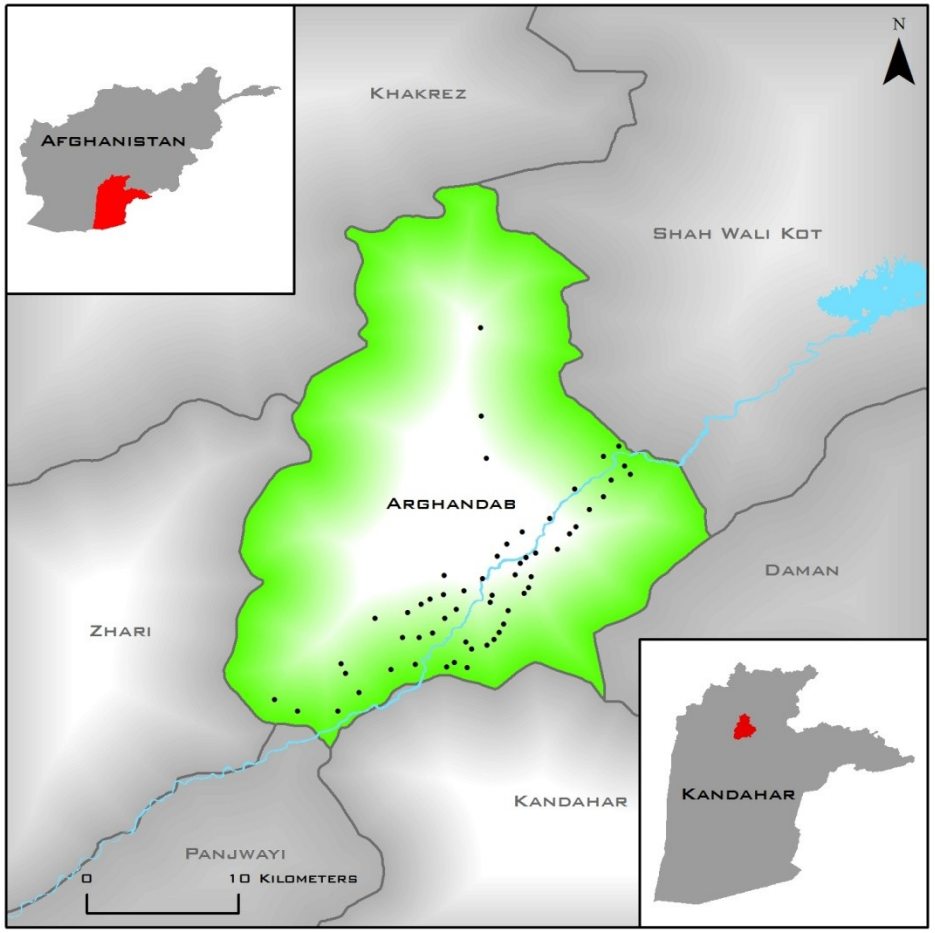

Figure 8. Arghandab district, Afghanistan showing the Arghandab river and village distribution.

Unit of Analysis

- Village level Thiessen polygons.
- Violent event counts aggregated at the village polygon level.
- Independent variables summarized at the village zonal unit (e.g. cumulative NDVI)
- *Assumes village influence/ownership of variables included within own zone.*

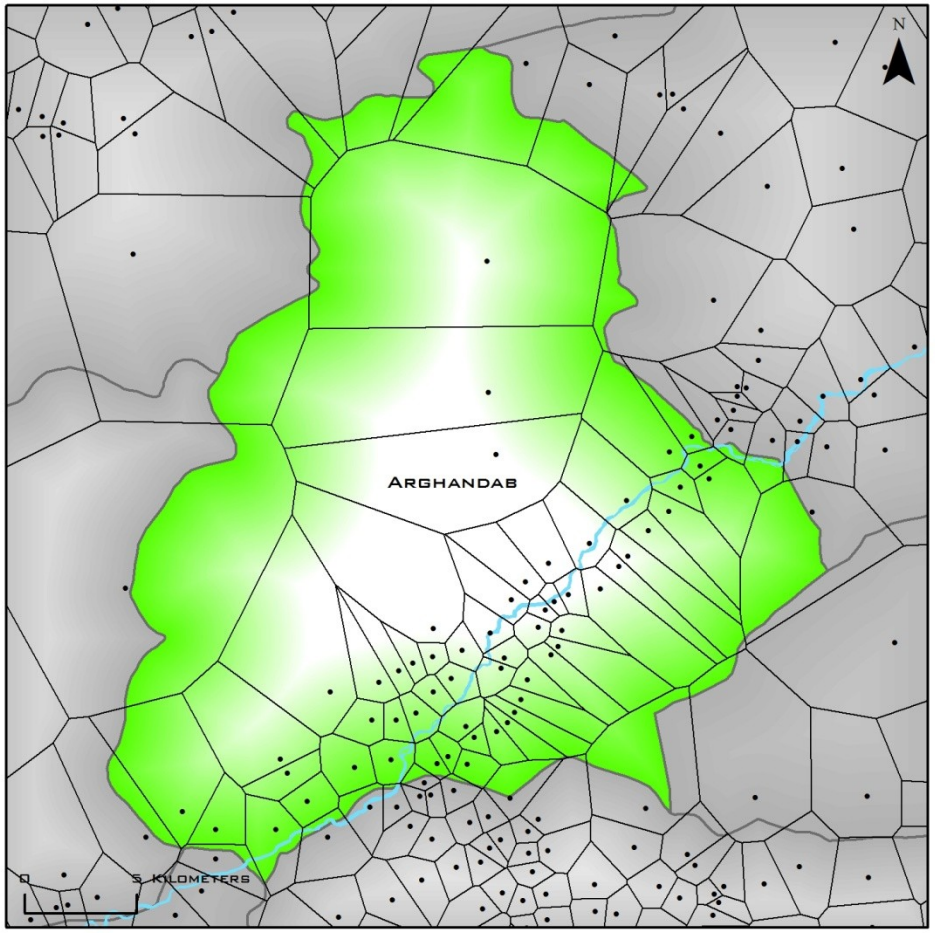

Figure 9. Study area and village level Thiessen polygon distribution.

Violent Events

- Obtained from a database of unclassified events since 2001.

- Subset of violent events including direct attacks, IED detonation/attempt/discovery, indirect attacks.

- Violent event counts taken at the village zone level and reported as a rate using residential land use as the base variable.

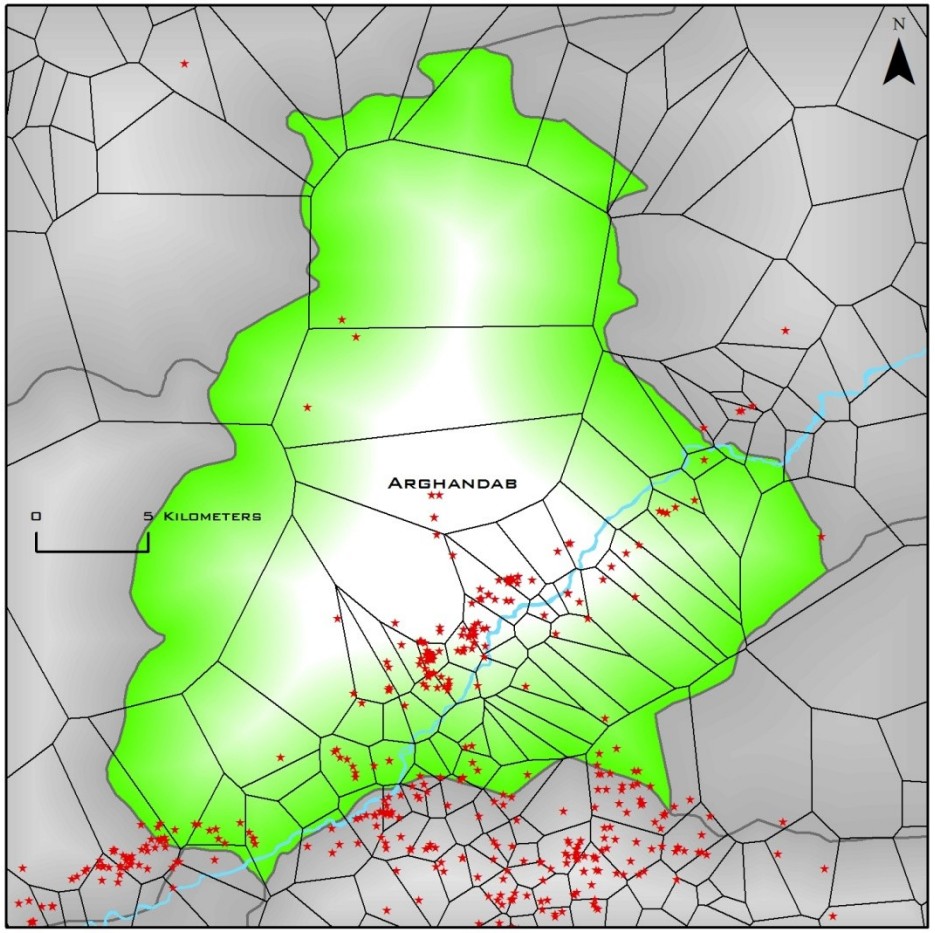

Figure 10. Study area showing a subset of violent event distribution.

Cumulative NDVI

- Derived from LandSAT 5/7 Enhanced Thematic Mapper
- Bimonthly coverage for each platform, study area lies within overlap, potential for 8 scenes per month.
- NDVI values reclassified to eliminate negative values (negative NPP).
- Zonal summaries of NDVI values for each village zone.
- *Assumes uniformity of crop selection.*

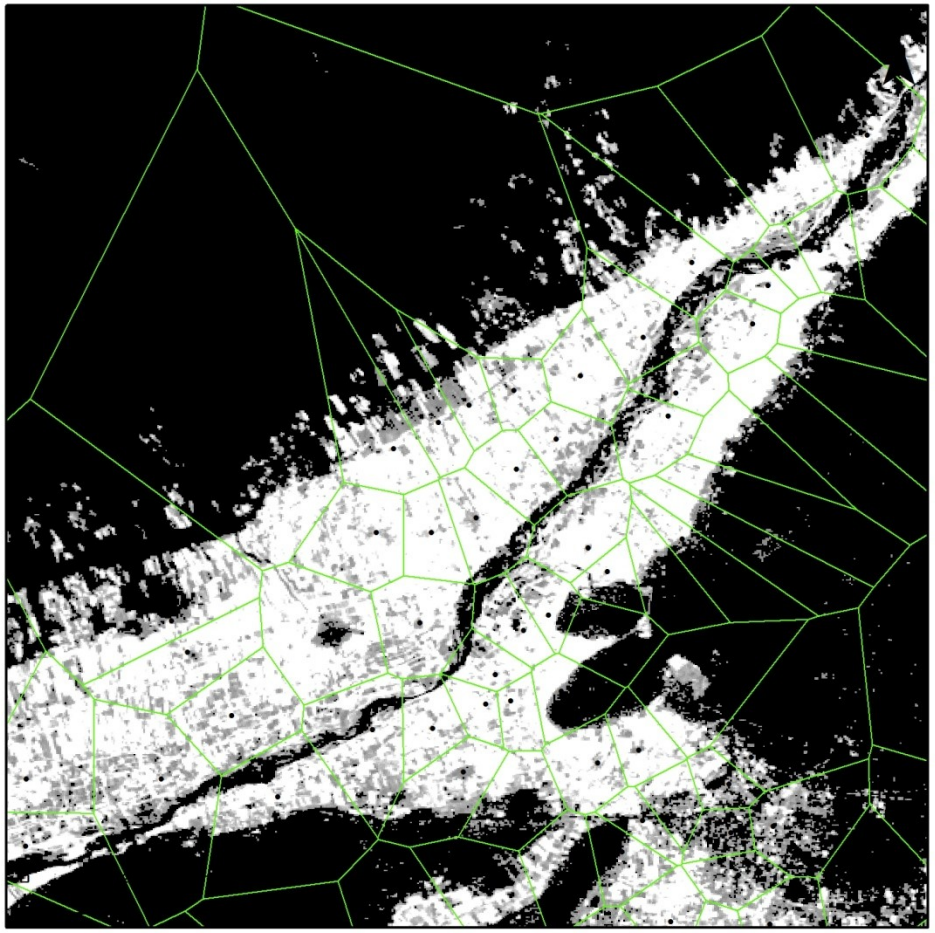

Figure 11. Sample NDVI calculation with village Thiessen overlay.

Residential Land use

- Derived from high resolution commercial imagery (Quickbird, IKONOS, Worldview).
- Residential land use extracted at 1:5000 on-screen.
- Residential land use area summarized for each village.
- Used as a proxy for population in the violent event rate calculation.
- *Assumes little variation in household size throughout the study area.*

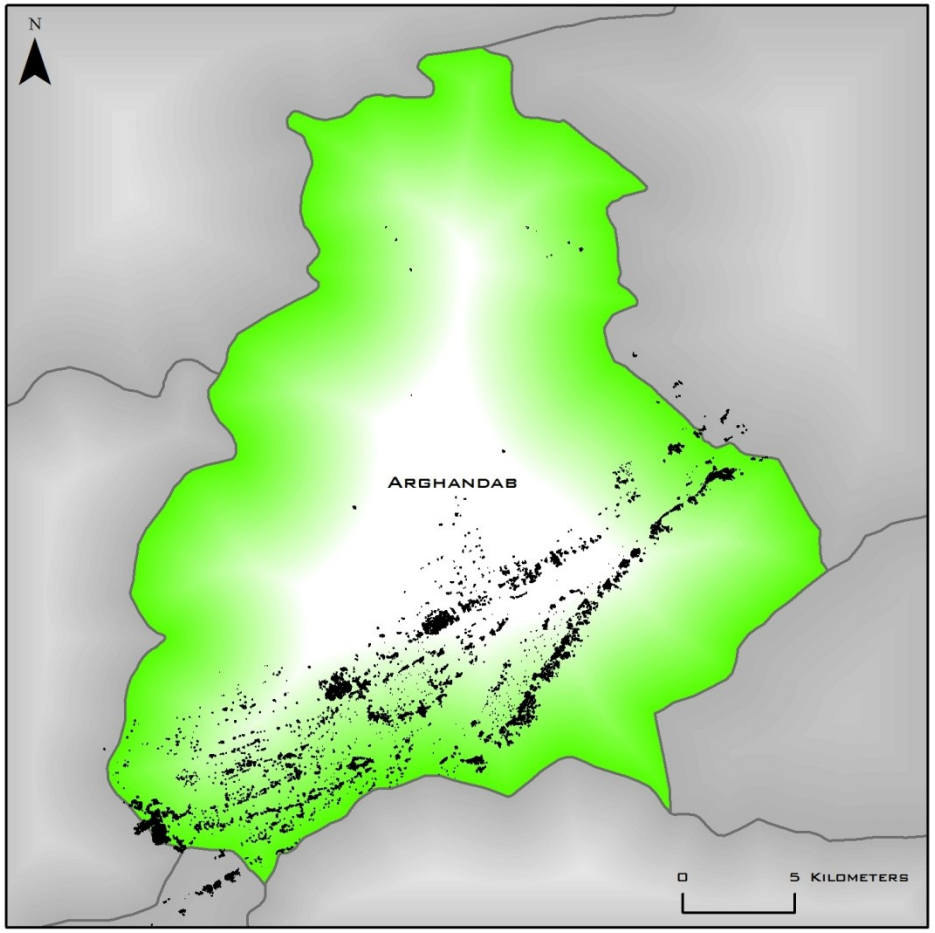

Figure 12. Study area showing residential land use distribution.

Local Indicators of Spatial Autocorrelation (LISA)

- Identifies statistically significant clusters of spatial autocorrelation.

- Used to identify areas where spatial autocorrelation in violent event rates and agricultural production may affect the OLS regression model.

- Conducted in OpenGeoDa to identify High-High and Low-Low clusters contributing to the Global Moran's I.

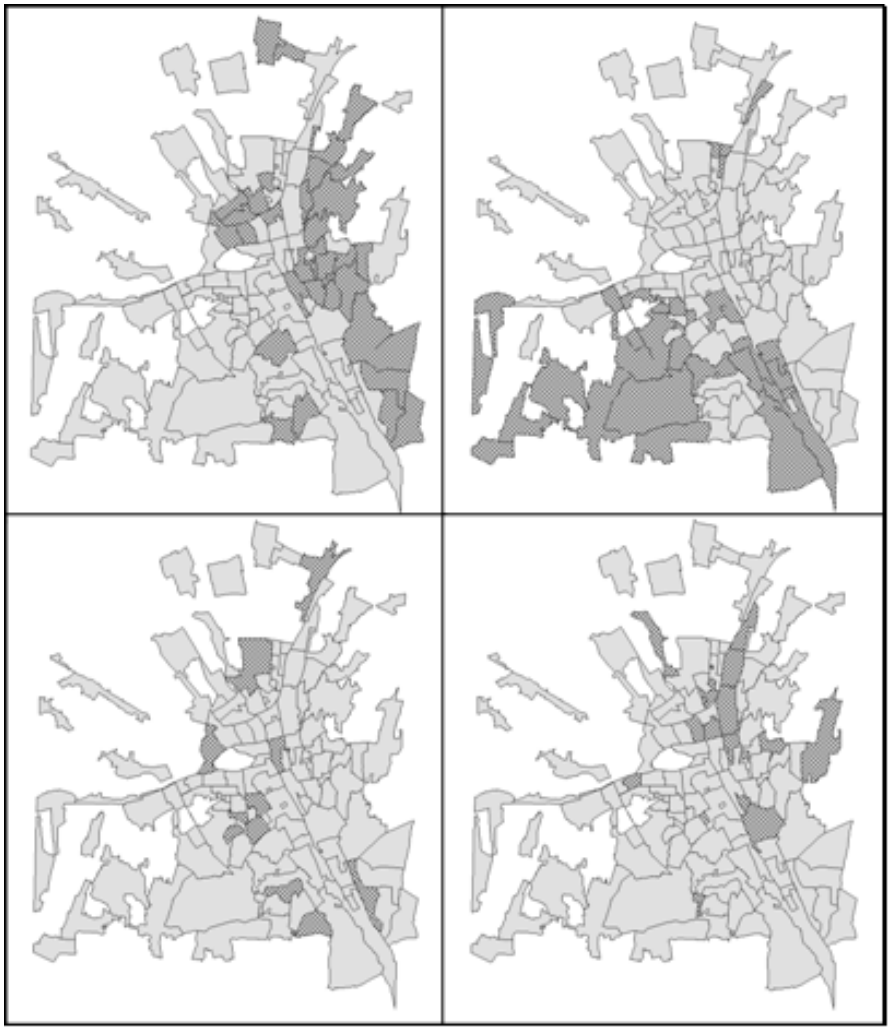

Figure 13. Sample OpenGeoDa LISA output.

Spatial Correlograms

- Plots univariate global Moran's I values at varying spatial lags.
- Assesses the extent of spatial autocorrelation in violent events and independent variables.
- Analogous to experimental semivariogram analysis in geostatistics.
- Conducted in R using the `sp.correlogram` function of the `spdep` package.

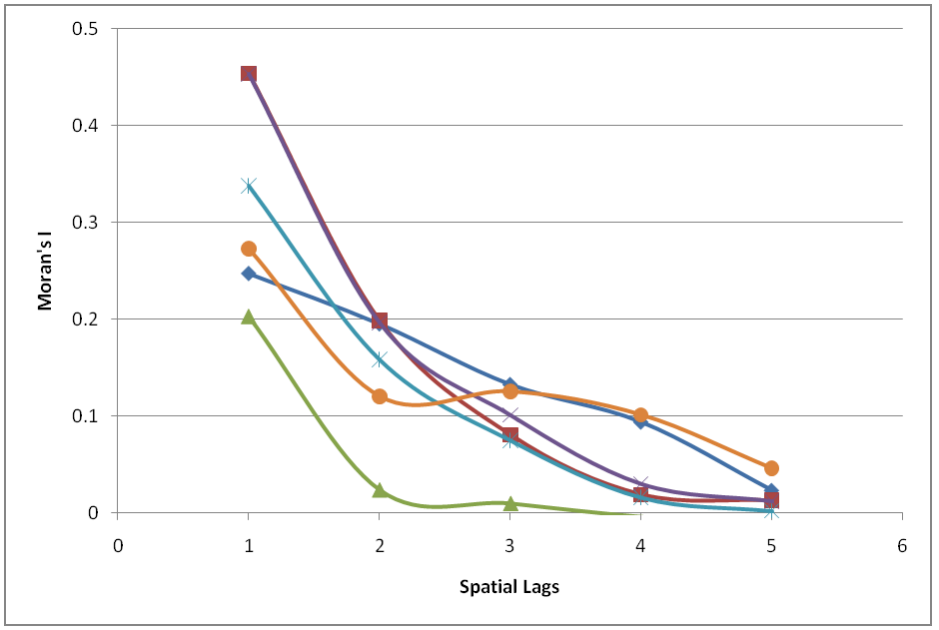

Figure 14. Sample spatial correlogram.

Spatial Regression Modeling

- OLS regression using cumulative NDVI, distance to nearest security force, Alikozai dummy variable, and distance to nearest road as explanatory variables.

- Assessment of variance and spatial dependence diagnostics in OpenGeoDa to determine if spatial lag or error model is needed.

- Varying weights matrices and independent variable observation month to model lag.

$$y = X\beta + \varepsilon$$

where:

y: n x 1 vector of violent event rates
X: n x 4 matrix of independent variables
ε: n x 1 vector of error terms

Lagged Regression Matrix

- Comparison of pseudo R^2 values at varying spatial and temporal lags.
- Total of 78 regression models used to identify stronger relationships at lag.
- Conducted up to six spatial lags away from the observations and does not include lower orders.
- Conducted at six months before and after the observations.

pseudo-R^2	W	$W^{(2)}$	$W^{(3)}$	$W^{(4)}$	$W^{(5)}$	$W^{(6)}$
t_{-6}	0.0062	0.0034	0.0059	0.0143	0.0062	0.0062
t_{-5}	0.0187	0.0276	0.0041	0.0109	0.0436	0.0945
t_{-4}	0.0265	0.0945	0.0223	0.0119	0.0499	0.0937
t_{-3}	0.0328	0.0828	0.0731	0.0127	0.1047	0.0439
t_{-2}	0.1381	0.1294	0.1987	0.1397	0.1143	0.1096
t_{-1}	0.2134	0.2345	0.2904	0.2232	0.1954	0.1545
t_0	0.2678	0.2712	0.3105	0.2956	0.2456	0.2213
t_1	0.4625	0.4659	0.5052	0.4903	0.4403	0.416
t_2	0.6572	**0.6606**	**0.6999**	**0.685**	0.635	0.6107
t_3	0.4749	0.4783	0.5176	0.5027	0.4527	0.4284
t_4	0.2926	0.296	0.3353	0.3204	0.2704	0.2461
t_5	0.1103	0.1137	0.153	0.1381	0.0881	0.0638
t_6	0.0668	0.0702	0.1095	0.0946	0.0446	0.0203

Figure 15. Sample lagged regression matrix.

Expectations and Concerns

- Higher pseudo R^2 values should be expected at 2-3 spatial lags and several months from the observed values.

- Potential for low or zero violent event counts in a large number of observations (right skewed).

- Possibility of little variation in model parameters (i.e. inability to identify superior model).

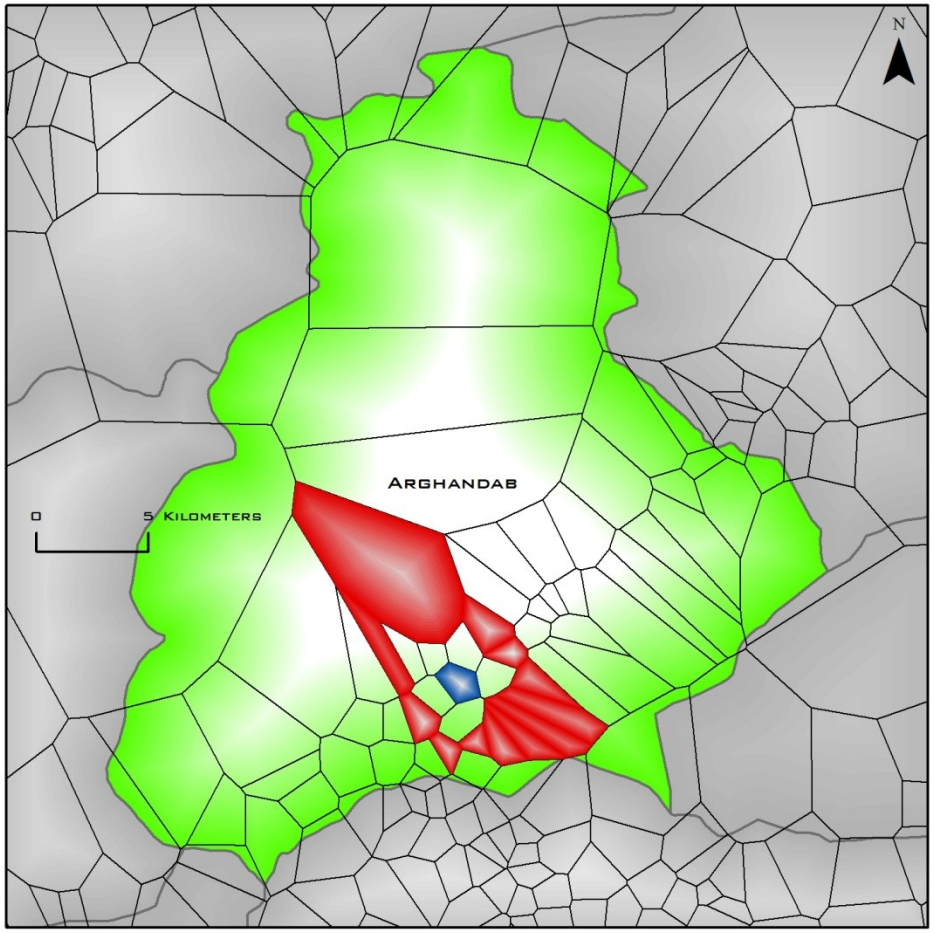

Figure 16. Theoretical relationship between DV (blue) and IV (red).

Potential Variations

- District level study of southern and southwestern districts.
- Substitute Afghan CSO district population figures for residential land use.
- Improved modeling of counterinsurgent security forces.
- Potential to identify inter-district attacks and lagged relationships (i.e. strategic concept).

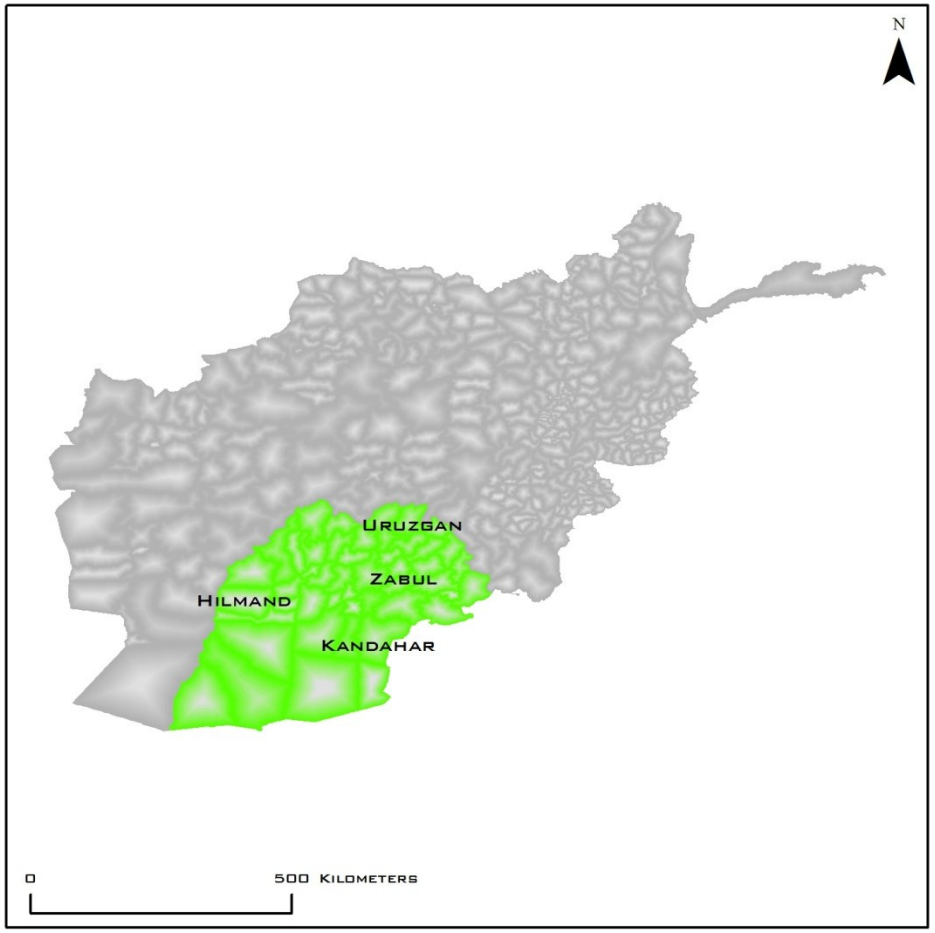

Figure 17. Possible district level approach using southern Afghanistan districts.

Questions or comments?