Examining Geospatial Covariate Relationships with Civilian Killings in South Sudan's Civil War

Rob Boswell

Introduction and Objective:

From Dec. 2013 - Feb. 2020, South Sudan experienced a civil war involving frequent targeting of civilians. A study by the London School of Hygiene & Tropical Medicine suggests as many as 400,000 civilians were killed. With the goal of identifying attributes to be used in machine learning models of mass killings in conflict-prone countries, I use the South Sudan case study to identify potential causal variables while accounting for spatial autocorrelation.

Theoretical Background & Methods:

Scholars posit various factors contribute to violence against civilians in conflict, including: ethnic groups' exclusion from state power, civilians living in regions along state borders where authorities lack presence, populations living far from capitals, mountainous terrain enabling rebels to hide and launch attacks (Collier and Hoeffler, 2004), droughts – which incentivize people to join rebel groups for income (Miguel et al., 2004), and combatants seizing territory where occupants are deemed loyal to the enemy (Straus, 2015). Perkoski and Chenoworth (2018) claim riots increase chances authorities will use violent repression. Other predictors include armed clashes between opponents, explosions and remote violence, and abductions.

From the ACLED dataset, violence against civilian events (green dots on the maps) and fatalities from such events (shades of red) are used as my dependent variables. Explanatory variables from ACLED include riots, territorial seizures, armed clashes, explosions and remote violence, and abductions. From the Prio-GRID dataset, distance from the capital, distance to the closest border, total rainfall, mountain terrain levels, and ethnic groups excluded from political power per cell grid (2012-13 only) were measured.

Steps:

1) Compare raster maps of civilian killings from 2012-2013 and 2014-2015 in terms of first order visual inspection, and **2)** maps of high and low clusters and global Moran's I statistics to verify the need for second order analysis via zonal statistics. **3)** Compare Lagrange multiplier diagnostics for spatial dependence to conclude which regressions are necessary. Examine and compare the results.

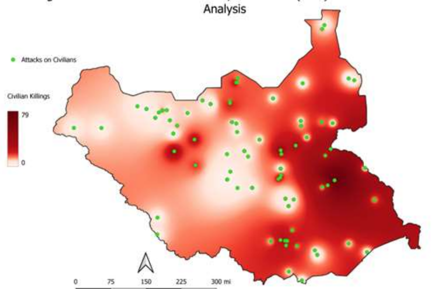
COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Note:

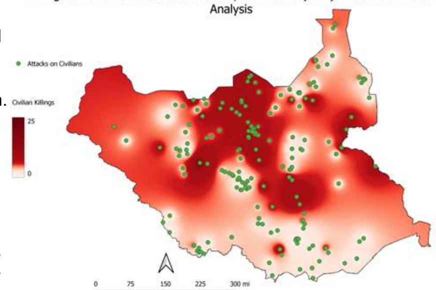
A first order Queen's contiguity matrix was found to have a higher Moran's I value than a second order Queen's for both 2012-13 and 2013-14.

Step 1:

Killings of Civilians Across South Sudan, 2012-2013 (IDW): First Order Visual Analysis

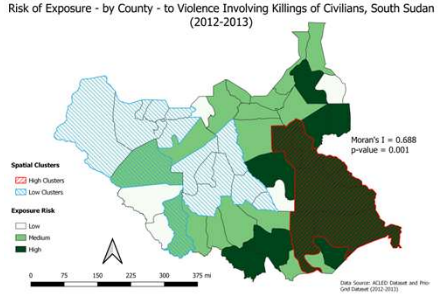


Killings of Civilians Across South Sudan, 2014-2015 (IDW): First Order Visual Analysis

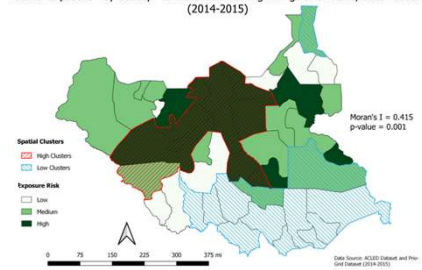


Step 2:

Risk of Exposure - by County - to Violence Involving Killings of Civilians, South Sudan (2012-2013)



Risk of Exposure - by County - to Violence Involving Killings of Civilians, South Sudan (2014-2015)



Step 3:

Spatial Dependence Diagnostics:

2012-2013

Test	MI/DF	Value	Prob
Moran's I (error)	0.3074	24.2317	0.00000
LM (lag)	1	218.7997	0.00000
Robust LM (lag)	1	76.589	0.00000
LM (error)	1	252.2818	0.00000
Robust LM(error)	1	110.0712	0.00000
LM (SARMA)	2	328.8708	0.00000

2014-2015

Test	MI/DF	Value	Prob
Moran's I (error)	0.1847	15.2872	0.00000
LM (lag)	1	101.1095	0.00000
Robust LM (lag)	1	25.8347	0.00000
LM (error)	1	91.1412	0.00000
Robust LM(error)	1	15.8665	0.00007
LM (SARMA)	2	116.976	0.00000

Brief Analysis:

In Step 1 (2012-13), it appears civilian fatalities were not randomly distributed. A disproportionate number of deaths appear in eastern and northern South Sudan, whereas cold spots appear in central and in northwestern South Sudan.

The pattern changed from 2014-15, whereby now civilian fatalities were high in the north (except for pockets of cold spots) and low civilian fatalities appeared in the south.

In Step 2, these patterns are largely confirmed. From 2012-13, a fairly strong Moran's I of 0.688 exists in which in the east, high local county civilian fatalities tend to exist alongside high levels of neighboring county fatalities. Likewise, in northwestern and central South Sudan, low local county civilian fatalities tend to coincide with low neighboring county fatalities. Similarly, the 2014-15 map shows High-Highs in the north, and Low-Lows in the south, confirming my first order analysis.

The Moran's I declines vis-à-vis 2012-13, but remains moderately strong at a statistically significant 0.415.

In Step 3, after running OLS models, both Moran's I (error) diagnostics for 2012-13 and 2014-15 show that model residuals are highly statistically significantly clustered in space. Further, in both cases, the LM lag and error diagnostics, and the Robust LM lag and error diagnostics are highly statistically significant. This indicates both spatial lag and error affect the OLS models. It is necessary to use a Spatial Durbin model, which controls for both spatial lag and spatial error clustering.

Spatial Durbin Results:

2012-2013

Column1	Estimate	Pr(>|z|)	Significant
(Intercept)	5.645	0.320	
mountains	-1.611	0.521	
distance to border	-0.021	0.003	***
distance to capital	-0.008	0.120	
1 group excluded	-0.035	0.966	
2 groups excluded	0.436	0.702	
3 groups excluded	-1.29	0.538	
rainfall amount	-0.001	0.557	
armed clashes	0.087	0.243	
explosions/remote violence	-0.374	0.007	***
riots	0.249	0.718	
territory seizures	2.743	0.001	***
lag.mountains	8.049	0.374	
lag.distance to border	-0.022	0.192	
lag.distance to capital	0.003	0.765	
lag.1 group excluded	-2.878	0.363	
lag.2 groups excluded	12.801	0.126	
lag.3 groups excluded	-47.005	0.002	
lag.rainfall amount	0.003	0.501	
lag.armed clashes	0.027	0.925	
lag.explosions/remote violence	0.329	0.493	
lag.riots	2.086	0.483	
lag.territory seizures	-1.904	0.507	

It is surprising that – once controlling for spatial lag and error – excluding groups from politics is not statistically significant. It is also surprising that distance to the capital is not significant. Yet, proximity to one's state border is highly statistically significant. And territorial seizure is also statistically significant. I thus interpret the coefficient as follows:

Each time a territory is seized by an armed group, there are – on average – 2.7 civilians who are killed, controlling for the other variables in the model, as well as for the lag of the error term for these other independent variables, and the for the overall spatial lag parameter (ρ).

```
Rho: 0.8948, LR test value: 60.078, p-value: 9.1038e-15
```

Further, ρ can be interpreted as: each additional civilian killed in neighboring counties in turn leads to about 0.89 civilians being killed in a given local county.

2014-2015:

Column1	Estimate	Pr(>|z|)	Significant
(Intercept)	4.788	0.058	
mountains	-0.124	0.869	
distance to border	0.006	0.002	***
distance to capital	0.003	0.031	**
rainfall amount	0.0004	0.541	
armed clashes	0.003	0.787	
explosions/remote violence	0.038	0.336	
riots	-0.321	0.0001	***
territory seizures	0.076	0.198	
lag.mountains	-6.595	0.051	
lag.distance to border	-0.003	0.587	
lag.distance to capital	-0.004	0.031	
lag.rainfall amount	-0.003	0.093	
lag.armed clashes	-0.041	0.122	
lag.explosions/remote violence	-0.310	0.009	
lag.riots	0.622	0.011	
lag.territory seizures	0.565	0.005	

It is notable that significant relationships between variables from 2012-2013 became insignificant from 2014-2015. This may result from the war's smaller death tolls from 2014-15 as seen in the raster layer legend.

Alternatively, it may well be that alternative fighting strategies were at play. This narrative could accompany the fact that the high-high and low-low clusters of violence against civilians switched dramatically in 2014 and 2015 compared to the prior two years.

It is also notable that the only relationship which remains significant (distance to nearest border) switches signs in 2014-15. This finding also suggests that the changing dynamic nature of the war made civilians previously safe by being further from the border at increasing risk of being killed.

The fact that the war became less brutal than before can be seen in the value of ρ, which declined. Now, for each individual killed by a neighboring county, one can expect 0.649 individuals killed in the local county.

```
Rho: 0.64871, LR test value: 18.113, p-value: 2.0817e-05
```

Conclusion:

Since some predictor coefficients lost or gained statistical significance over time, it may be that there are conditional relationships at play not easily discernible. Further, all predictors should still be used in machine learning models. As Ward et al. (2010) explain, at times even variables with high p-values end up being good predictors, though not necessarily causal.