A Statistical comparison between ECAV and SCAD event data

Elio Amicarelli

doc ref: ECAV03017

1. Analysis and results

This report presents a statistical comparison between the beta version of the Electoral Contention and Violence (ECAV) and the Social Conflict Analysis Database (SCAD). It examines country level as well as temporal and spatial characteristics of the two datasets. The sample includes 41 African countries available in both ECAV and SCAD.[1] Since the two datasets differ in their scope as well as in their unit of observation, SCAD was preprocessed by expanding each event according with its duration and by filtering out the events not having "elections" listed among their issues. Except for the spatial analysis, only observation with "sublocal" equals to 1 were retained in order to exclude duplicate events. SCAD does not explicitly link events to specific elections, for this reason it was not possible to use country-election as unit for the analyses. Instead, the list of electoral periods as defined by the ECAV projects were used to filter out events falling outside of these periods.

Country-level analysis

Table 1 (p.4) shows the number of events by country. In 29 cases ECAV contains more events than SCAD and in total it has approximately 2000 more events. The number of events by country is plotted in *Figure 1* (p.5) together with a blue line of theoretical perfect agreement. In both datasets the majority of countries are clustered in the range between 0 and 200 events. Algeria, Nigeria, Egypt, Kenya and Zimbabwe represent notable exceptions. With 25 events in SCAD and 613 events in ECAV, Algeria stands as a suspicious case of marked disagreement between the two datasets. The country-level correlation between the total number of events contained in ECAV and SCAD is 0.67 (95% CIs: 0.4517163 0.8085236). *Figure 2* (p.5) shows how the correlation coefficient on the y-axis is affected by the removal of individual countries. The correlation coefficient is positively affected by the individual removal of Algeria and Zimbabwe while the removal of Egypt seems to decrease the overall correlation. In particular, the removal of Algeria or Zimbabwe gives a correlation of 0.73 (95% CIs: 0.5565492, 0.8544068) and 0.80 (95% Cis: 0.6521556 0.8903211) respectively. The influential nature of Algeria, Egypt, Kenya and Zimbabwe on the agreement between the two datasets is also confirmed by *Figure 3* (p.6) which plots the Cook's distance and the dfbeta resulting from a regression of ECAV on SCAD (left panel) and from a regression of SCAD on ECAV (right panel). The correlation between ECAV and SCAD is 0.69 (95% CIs: 0.4552151 0.8221841) when these four influential observations are collectively removed from the sample. The plot in *Figure 4* (p.6) compares the total number of events in ECAV and SCAD for countries with less than 400 events (the blue line is the theoretical perfect agreement). As can be seen from the plot, this particular area of the data presents considerable variation and a general tendency for ECAV to have more events than SCAD. The correlation in this portion of the data is 0.40 (95% CIs: 0.07994741 0.64254144). *Figure 5* (p.7) shows that the correlation is relatively robust to the removal of individual countries.

Temporal analysis

For the temporal analysis, the total number of events was calculated relying on country-month as unit for the calculation. Only countries included in both ECAV and SCAD were considered. The months to be considered

---

[1] Angola was removed because not contained in ECAV while Botswana, Comoros and Djibouti were removed because not contained in SCAD.

for the analysis were obtained from the table defining the temporal windows used to retrieve the material for the ECAV data collection. Both datasets were augmented with months that they originally not included because of absence of events. These months were assigned a value of 0 as total number of events. Using these country-month observations, the overall correlation between ECAV and SCAD is 0.45 (N = 2260; 95% CIs: 0.4216483 0.4870656) however, *Figure 6* (p.7) shows that if we look at the temporal correlation for each country, it ranges from very low to very high levels. We can examine the monthly-based correlation for those countries previously identified as particularly interesting cases: Algeria, Egypt, Kenya and Zimbabwe. At the country-month level, SCAD and ECAV correlate very poorly on Algeria, moderately well on Egypt, and well on Kenya and Zimbabwe. This provides additional evidences for the particularly suspicious nature of Algeria. Indeed, the first panel in *Figure 7* (p.8) highlights that the difference in the ECAV and SCAD time-series for Algeria is striking: according with ECAV the country has a very dynamic history of electoral contention and violence, while according with SCAD it seems a largely peaceful case. ECAV also presents a more intense activity than SCAD for Egypt, showing generally higher and more frequent spikes in the number of events. The series for Kenya has a clear maximum peak toward its end in ECAV but not in SCAD. ECAV and SCAD seem to have fair degrees of agreement on Zimbabwe, especially in regard with when events are observed and when they are not. However, there are differences regarding the total number of events observed in each month with SCAD generally recording more activities. More specifically, it seems that when a relatively high number of events is recorded in both datasets, SCAD generates higher spikes. This phenomenon could be generated by the fact that SCAD events duration were used to expand the dataset for this analysis. It is important thus to remember when looking at this result that ECAV and SCAD differ in their definition of "event". While ECAV codes an event as a specific episode taking place on the ground, SCAD adopts a broader definition and codes an event as an episode lasting for a period with a start and end date. When expanding SCAD events based on their duration, we assume that events' activities took place on the ground during the entire duration. This assumption can lead to inflate the actual level of activity during the period of interest. Further analyses are required to verify this conjecture. The remaining panels of *Figure 7* show time-series with other interesting behaviors. For example, Tanzania and Ethiopia have a high number of events in SCAD around periods 26 and 36 respectively while these peaks are absent in ECAV. Since it is not possible to go through all 41 countries in this report, an interactive web application was developed and deployed at the following link: https://elioa.shinyapps.io/interface1/. The application allows comparing time series in ECAV and SCAD at the monthly and weekly level, and identifying the temporal coordinates of all the data points in the series. This tool can be used to strategically identify countries and temporal segments that might be worth a closer look for subsequent investigations.

Spatial analysis

Both ECAV and SCAD are georeferenced datasets. The spatial analysis conducted examines the geographical dispersion of the events in the two datasets as well as the proximities of their locations. Initially, each country was divided in cells of 20 squared kilometers. *Table 2* (p.9) shows i) the total number of cells, ii) the total number of cells with at least one ECAV event, iii) the total number of cells with at least one SCAD event, and iv) how many cells have both at least one ECAV event and one SCAD event. Except for Libya, the number of unique cells active in ECAV is always higher than the respective number in SCAD. This tells us that ECAV events have higher geographical variation than SCAD events. At first sight the intersection between ECAV and SCAD active cells does not seem very high. There are two possible reasons for this low intersection. First, it could be the case that part of SCAD and ECAV events are taking place in distant areas of a country; or second, the activities recorded in the two datasets are near but their locations or coordinates are not exactly the same. In an ideal scenario with full intersection the ratio between values of the "intersection" column and values of the "SCAD" column from Table 2 would be equal to 1 for all countries. *Figure 8* (p.10) shows how this ratio changes as we increase the size of the cells. As can be seen, by increasing the size of the cells the density moves toward higher levels of intersection. This means that ECAV events tend to be present in the same areas populated by SCAD events even though the geographical coordinates of the specific events are not exactly the same. Additional analyses on ECAV geolocation precision are strongly encouraged. *Figure 9* (p.10) provides a visual example for the higher geographical variation in ECAV's locations as well as for the size of the 20 km cells.

Violence and contention

The final part of this report examines some aspect of electoral violence and contention. *Table 3* (p.11) shows the total number of deaths by country in SCAD and ECAV. The total number of deaths for SCAD were calculated before expanding the events by their duration in order to not inflate the number of casualties. Since ECAV adopts an interval scale to define the number of casualties associated with an event, Table 3 specifies a minimum and maximum estimate for each country. Since the last category in ECAV has no upper bound, the category's lower bound was doubled and used as upper bound. In 20 cases the total number of deaths recorded by SCAD is within the range calculated for ECAV. However, there are important differences for Algeria, Congo, Ghana, Libya and Sudan. As already occurred in previous sections of this analysis, SCAD and ECAV provides two very different pictures on the situation in Algeria. According with ECAV, Algeria is the deadliest country during electoral periods while SCAD is not recording casualties at all. *Table 4* (p.12) concludes the analysis by showing the proportions of violent and nonviolent event by country for the two datasets. A Chi-Square test for the equality of proportions was performed and the results are reported in the last column.[2] According with this test, the null hypothesis of equal proportions is rejected for all countries except for very few cases, namely Benin, DRC, Egypt and Mauritania.

2. Concluding remarks

This report compared ECAV and SCAD data on different dimensions of interest. The two datasets reach moderate levels of correlation in regard with the total number of events per country. The temporal analysis revealed considerable variation in respect with the level of agreement between country-level time-series and highlighted general aspects as well as interesting cases of disagreement. The analysis of the spatial distribution of events showed that ECAV events take place in a wider range of locations than SCAD. In addition, even if SCAD events tend to be geolocated in areas populated by ECAV events, in several cases their exact geographical coordinates differ. In the final part, the analysis showed that the total number of deaths by country recorded by SCAD is often within the range estimated by ECAV. It also provided preliminary evidences for the two datasets to significantly differ in regard with the countries' proportions of violent and nonviolent events. This report showed that ECAV contains more events than SCAD and that it is characterized by more nuanced temporal and spatial variation. The datasets differ under several aspects and a natural extension of this analysis would consist in closely examining a set of specific cases. This report highlighted some critical cases of disagreement such as Algeria and provided a web tool to support further interactive examinations of the data. The web tool allows to easily spot the temporal periods where the ECAV and SCAD time series markedly differ. It can be used to strategically select a set of countries where the two datasets differ or align in interesting ways and to support future analyses aiming at tracing differences in the specific events recorded by the two datasets.

---

[2]The null hypothesis for this test is that the proportions are equal.

3. Figures and tables

3.1 Country-level analysis

Country	ECAV	SCAD
Algeria	613	25
Benin	22	19
Burkina Faso	5	10
Burundi	90	23
Cameroon	31	111
Central African Republic	72	8
Chad	15	3
Congo	74	57
Cote D'Ivoire	161	84
Democratic Republic of Congo	121	43
Egypt	1053	521
Ethiopia	69	45
Gabon	75	49
Gambia	4	4
Ghana	33	54
Guinea	61	97
Guinea-Bissau	34	7
Kenya	619	164
Lesotho	12	6
Liberia	7	35
Libya	20	15
Madagascar	95	173
Malawi	33	100
Mali	34	90
Mauritania	23	13
Mauritius	4	8
Morocco	23	3
Mozambique	41	1
Namibia	7	1
Niger	57	15
Nigeria	491	224
Rwanda	32	7
Senegal	65	25
Sierra Leone	25	42
South Africa	212	85
Sudan	113	19
Tanzania	91	94
Togo	77	46
Uganda	134	15
Zambia	54	47
Zimbabwe	434	777
TOTAL	5236	3165

Table 1: Total number of events by country

Events by country

Figure 1

Correlation after single case deletion

Figure 2

Regress ECAV on SCAD
Influential data points

Regress SCAD on ECAV
Influential data points

Figure 3

Total events by Country

Figure 4

Correlation after single case deletion
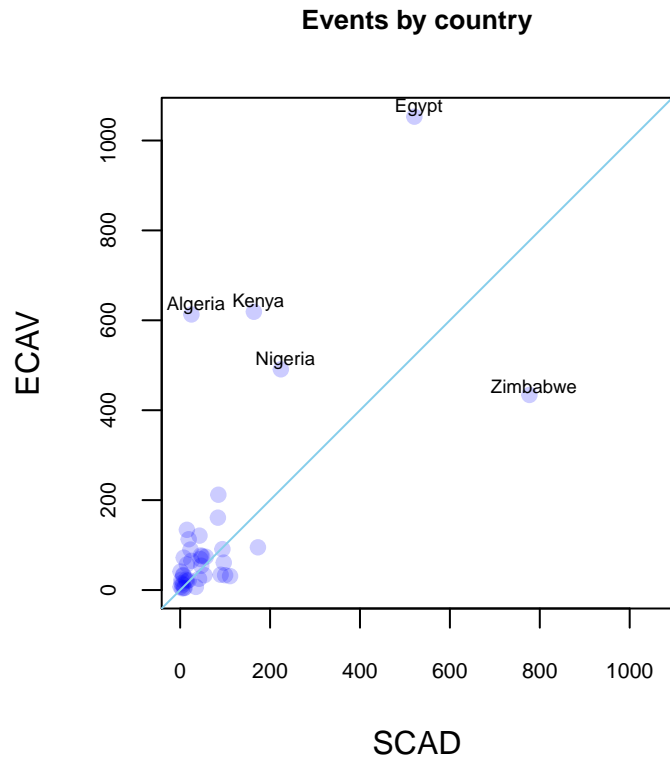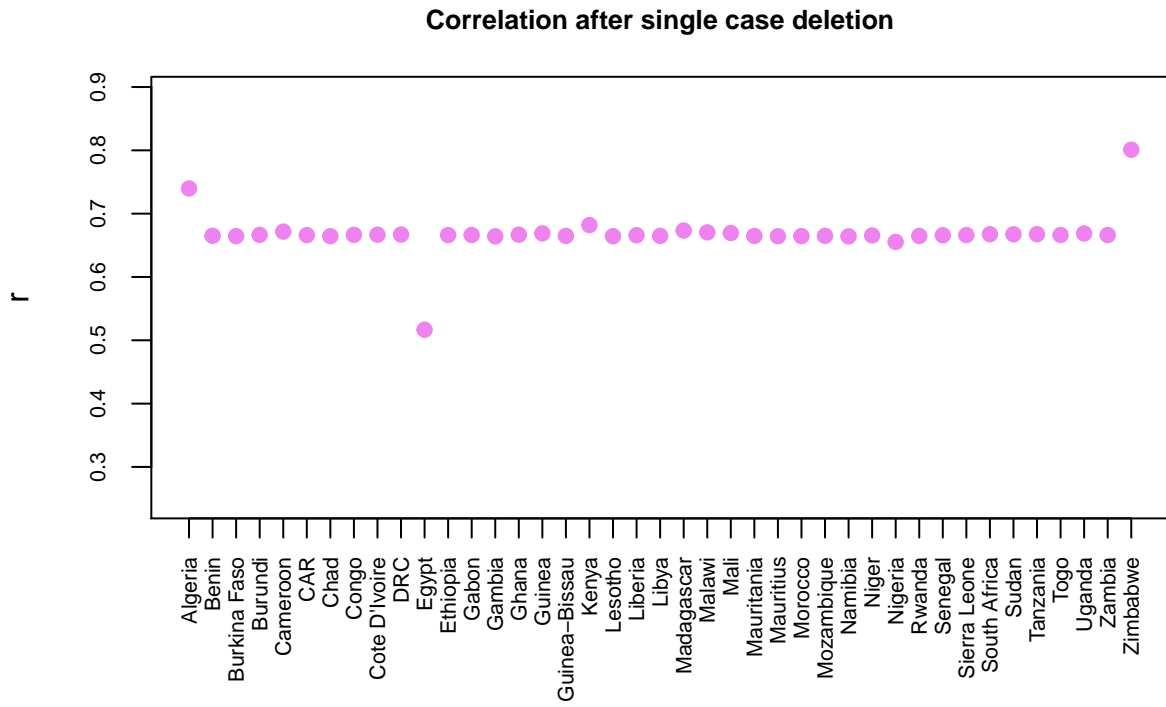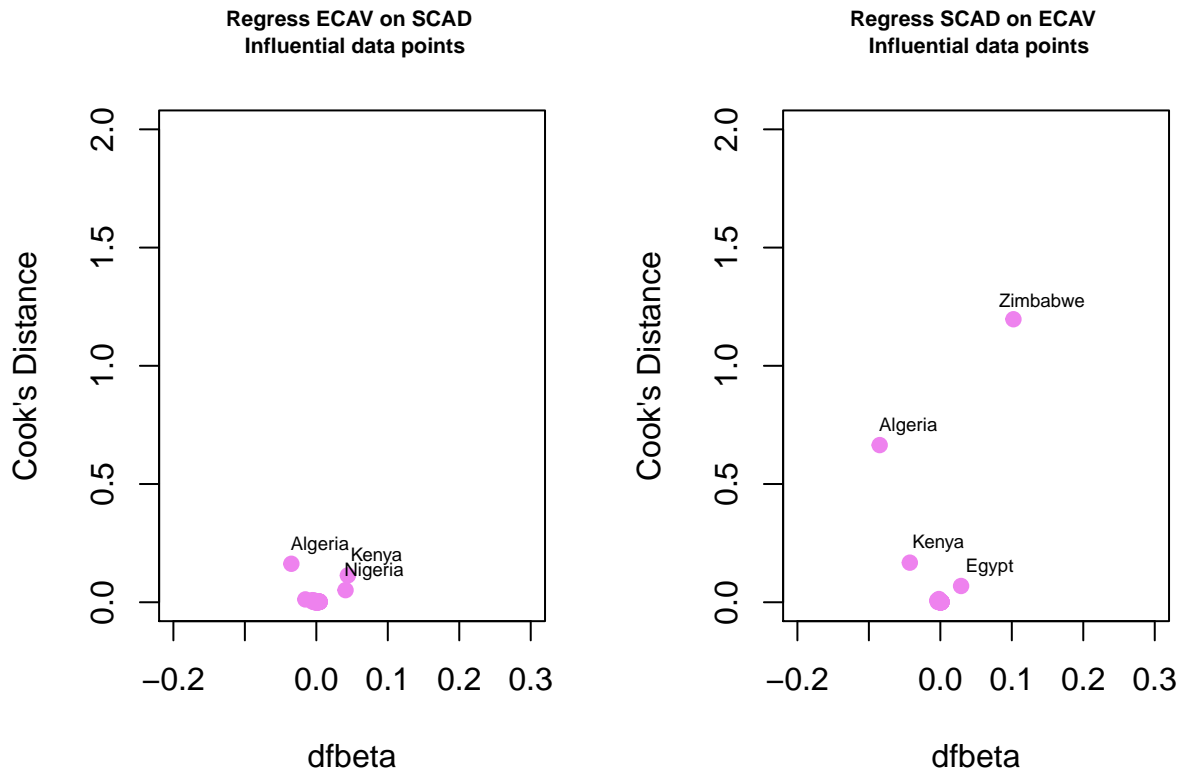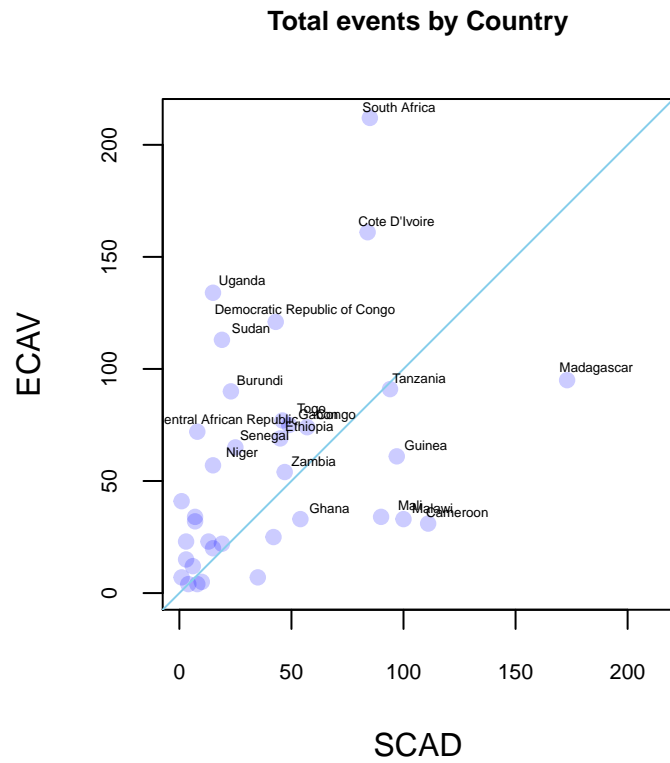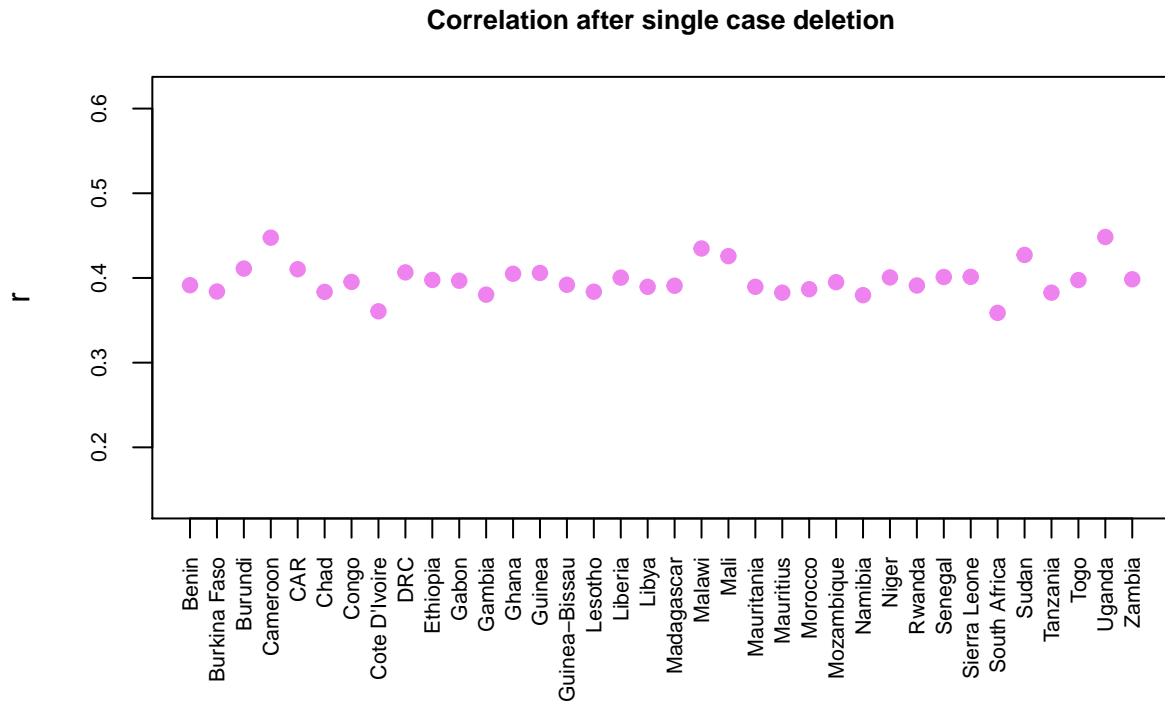
Figure 5

3.2 Temporal analysis

Correlation at the country−month level

Figure 6

Algeria

Egypt

Kenya

Zimbabwe

Madagascar

Uganda

Tanzania

Ethiopia
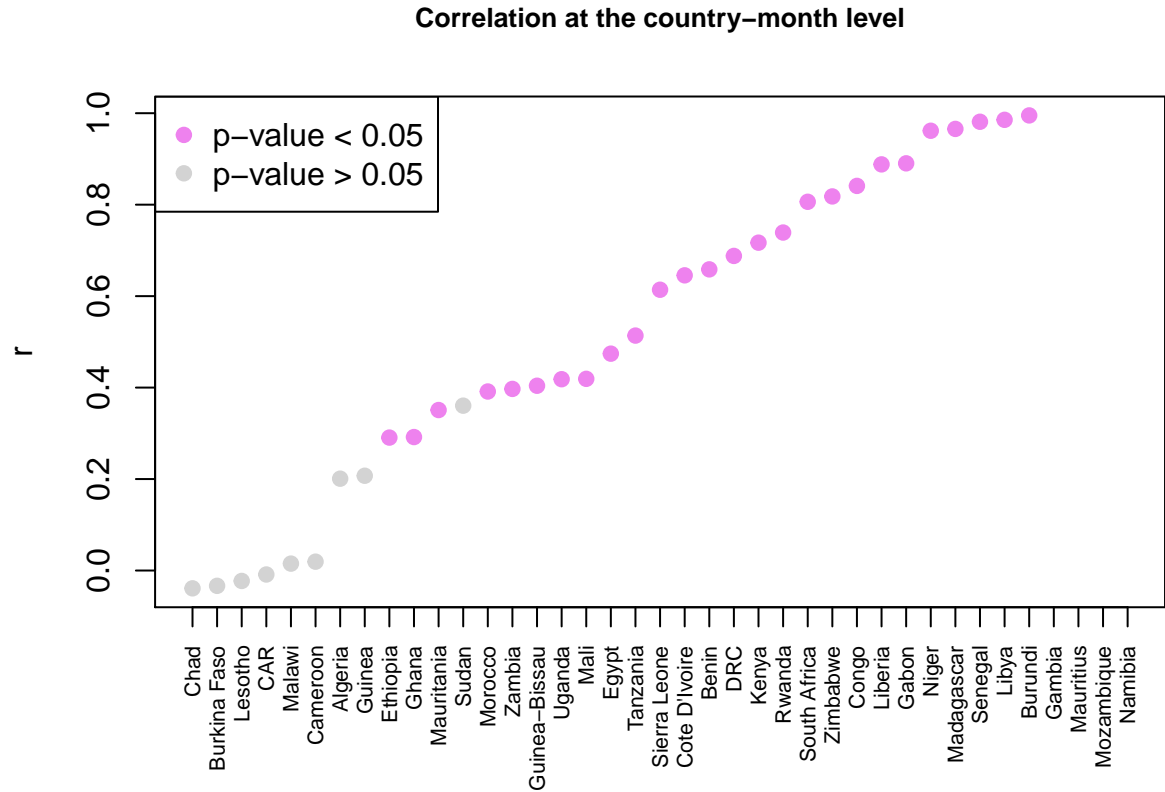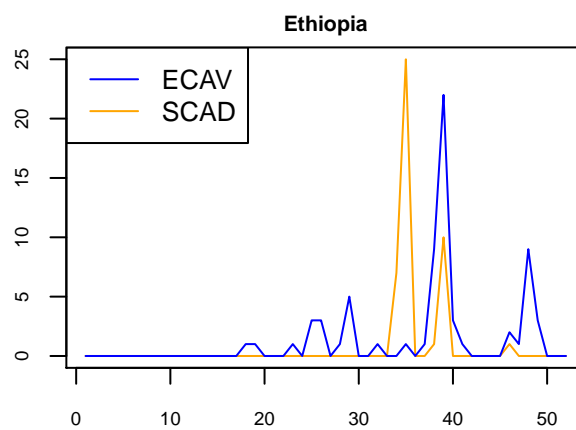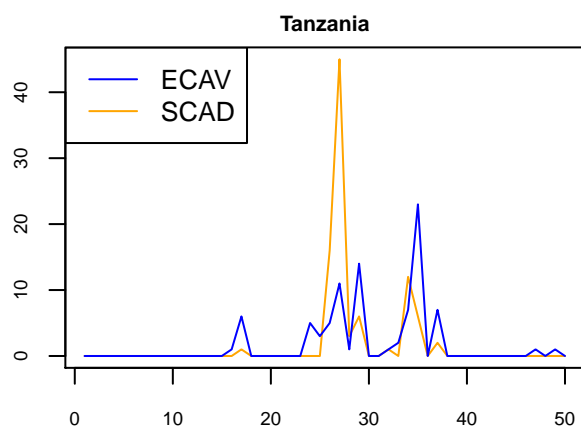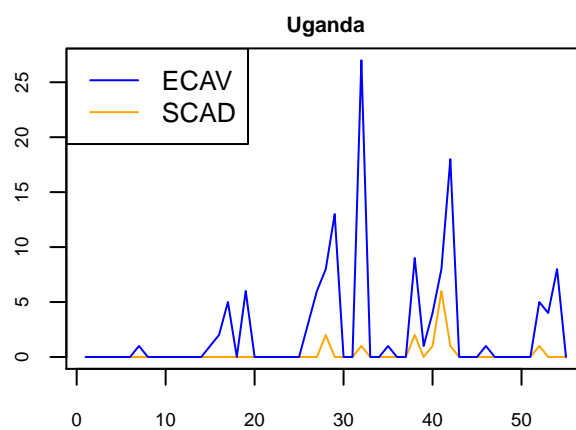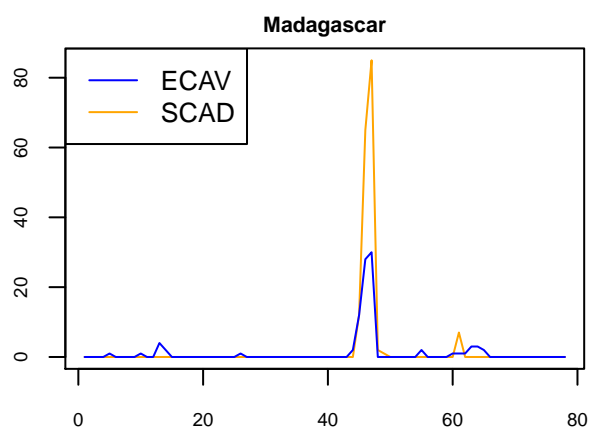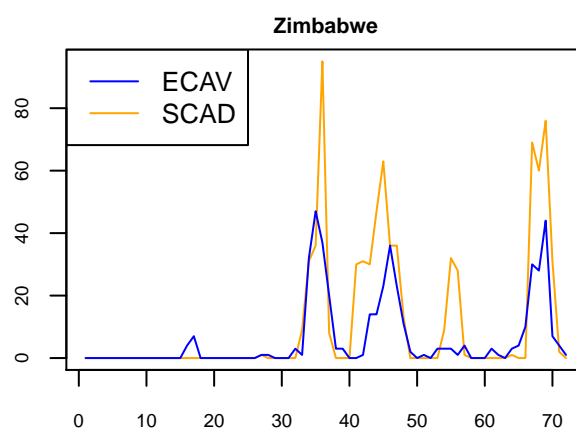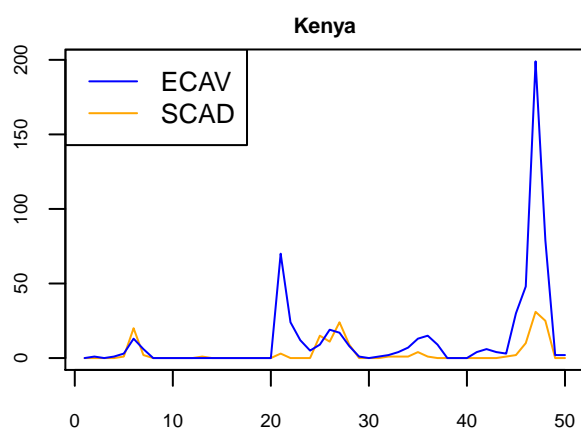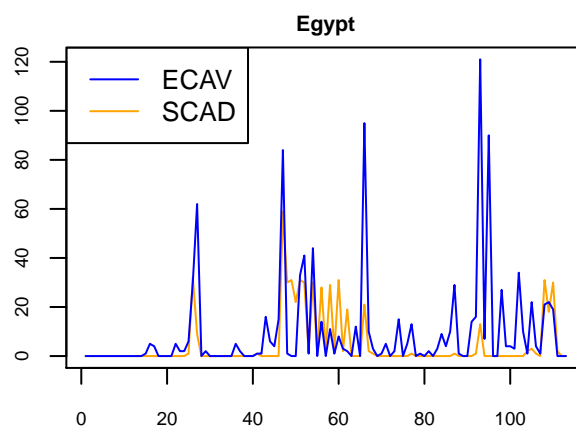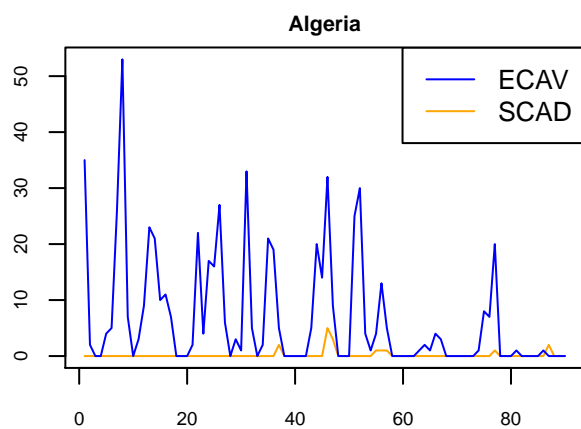
Figure 7

3.3 Spatial analysis

Country	N	ECAV	SCAD	intersection
Algeria	5536	99	5	2
Benin	280	7	6	2
Burkina Faso	644	2	1	0
Burundi	71	12	3	2
Cameroon	1067	13	9	5
Central African Republic	1388	7	6	3
Chad	2820	3	2	1
Congo	815	4	2	0
Cote D'Ivoire	735	13	5	3
Democratic Republic of Congo	4991	27	15	9
Egypt	2372	74	13	7
Ethiopia	2458	20	4	2
Gabon	602	6	9	3
Gambia	40	1	1	1
Ghana	552	14	11	4
Guinea	588	10	3	1
Guinea-Bissau	103	7	2	0
Kenya	1295	110	26	21
Lesotho	90	1	1	1
Liberia	234	4	3	2
Libya	3867	6	9	1
Madagascar	1395	13	6	3
Malawi	316	10	6	2
Mali	2838	7	3	1
Mauritania	2382	3	3	0
Mauritius	13	2	1	1
Morocco	1564	9	2	0
Mozambique	1861	29	2	2
Namibia	1944	5	1	1
Niger	2667	7	5	1
Nigeria	1998	107	27	16
Rwanda	67	6	2	1
Senegal	493	9	8	1
Sierra Leone	181	8	7	4
South Africa	3039	65	32	16
Sudan	4103	16	3	2
Tanzania	2062	18	3	2
Togo	152	11	4	3
Uganda	552	41	3	2
Zambia	1702	9	5	4
Zimbabwe	910	79	11	5
TOTAL	60787	894	270	137

Table 2: Unique cells (20km) occupied by ECAV and SCAD events

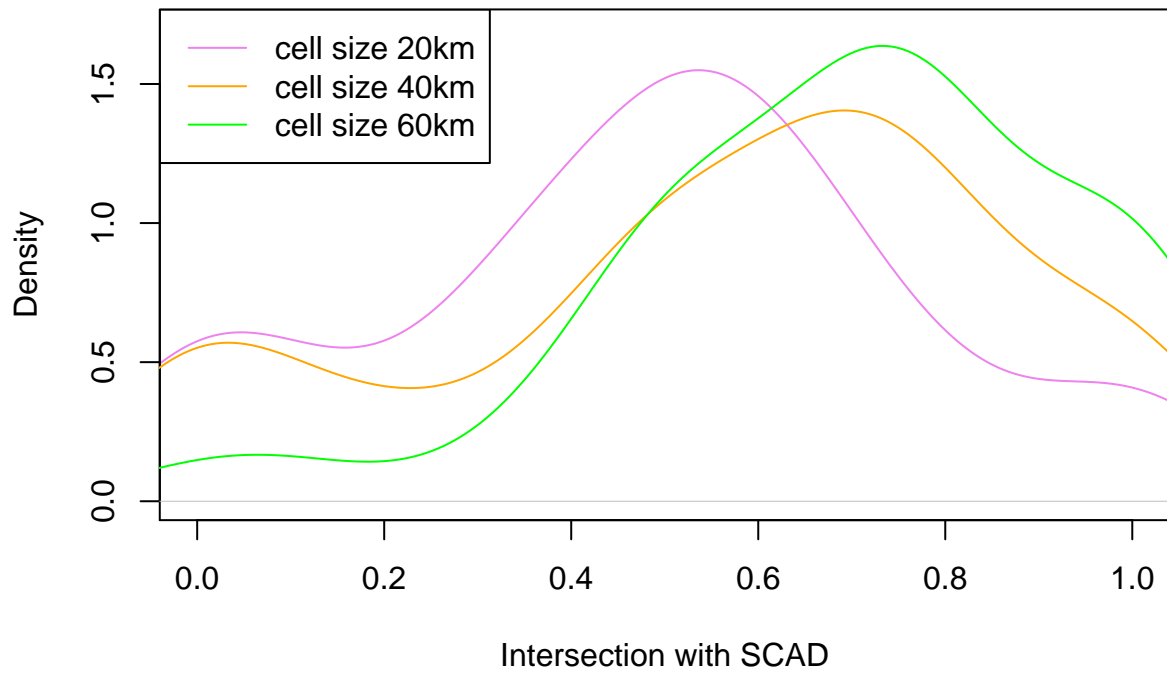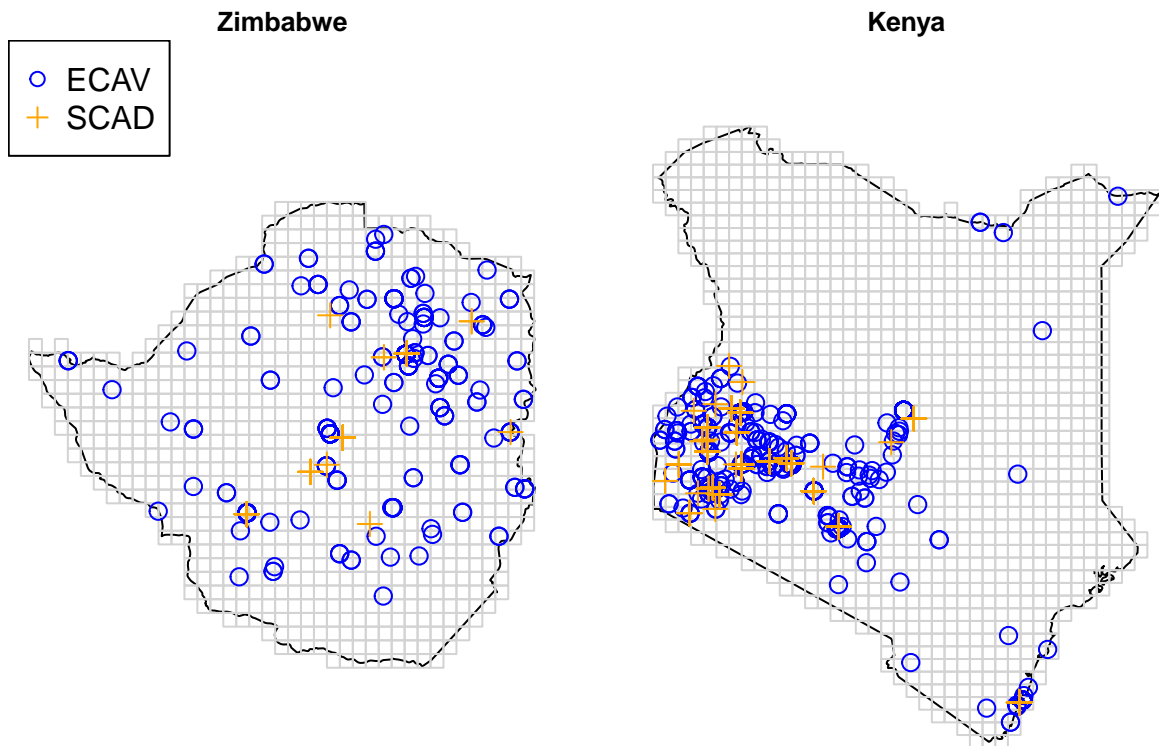## Active cells' intersection as share of SCAD active cells

Figure 8

Figure 9

3.4 Violence and contention

Country	SCAD	ECAV(min-max)
Algeria	0	937-8291
Benin	4	2-18
Burkina Faso	0	2-18
Burundi	16	12-108
Cameroon	7	5-45
Central African Republic	15	5-45
Chad	1	3-27
Congo	1	139-569
Cote D'Ivoire	469	175-1665
Democratic Republic of Congo	110	59-567
Egypt	82	152-1422
Ethiopia	53	36-342
Gabon	8	3-27
Gambia	2	2-18
Ghana	20	105-245
Guinea	68	28-261
Guinea-Bissau	3	10-90
Kenya	1744	556-4529
Lesotho	0	2-18
Liberia	0	1-9
Libya	2	53-522
Madagascar	26	12-108
Malawi	51	5-45
Mali	3	0-0
Mauritania	7	1-9
Mauritius	1	1-9
Morocco	0	5-45
Mozambique	0	0-0
Namibia	0	10-99
Niger	0	3-27
Nigeria	967	139-1305
Rwanda	6	6-54
Senegal	11	52-504
Sierra Leone	28	41-405
South Africa	163	194-1854
Sudan	45	325-1615
Tanzania	44	28-270
Togo	111	25-234
Uganda	16	33-306
Zambia	2	4-36
Zimbabwe	118	89-819

Table 3: Total number of deaths by country

Country	ECAV violent	SCAD violent	ECAV nonviolent	SCAD nonviolent	Chi-Square
Algeria	0.64	0.44	0.36	0.56	4.03 *
Benin	0.59	0.47	0.41	0.53	0.56
Burkina Faso	0.60	0.00	0.40	1.00	.
Burundi	0.84	0.87	0.16	0.13	.
Cameroon	0.55	0.16	0.45	0.84	19.46 ***
Central African Republic	0.12	1.00	0.88	0.00	.
Chad	0.47	0.33	0.53	0.67	.
Congo	0.73	0.05	0.27	0.95	60.06 ***
Cote D'Ivoire	0.84	0.63	0.16	0.37	14.31 ***
Democratic Republic of Congo	0.74	0.81	0.26	0.19	1.06
Egypt	0.46	0.43	0.54	0.57	2.01
Ethiopia	0.36	0.73	0.64	0.27	15 ***
Gabon	0.41	0.18	0.59	0.82	7.15 **
Gambia	0.75	1.00	0.25	0.00	.
Ghana	0.76	0.91	0.24	0.09	.
Guinea	0.64	0.08	0.36	0.92	55.57 ***
Guinea-Bissau	0.56	0.57	0.44	0.43	.
Kenya	0.88	0.95	0.12	0.05	6.45 *
Lesotho	0.58	0.17	0.42	0.83	.
Liberia	0.57	0.00	0.43	1.00	.
Libya	0.90	0.67	0.10	0.33	.
Madagascar	0.36	0.06	0.64	0.94	40.24 ***
Malawi	0.82	0.94	0.18	0.06	.
Mali	0.82	0.19	0.18	0.81	42.99 ***
Mauritania	0.52	0.62	0.48	0.38	0.29
Mauritius	0.25	1.00	0.75	0.00	.
Morocco	0.52	0.00	0.48	1.00	.
Mozambique	0.78	1.00	0.22	0.00	.
Namibia	0.57	0.00	0.43	1.00	.
Niger	0.58	0.00	0.42	1.00	16.03 ***
Nigeria	0.58	0.81	0.42	0.19	36.24 ***
Rwanda	0.38	0.71	0.62	0.29	.
Senegal	0.72	0.36	0.28	0.64	10.13 **
Sierra Leone	0.92	0.67	0.08	0.33	5.53 *
South Africa	0.80	0.34	0.20	0.66	56.77 ***
Sudan	0.69	0.26	0.31	0.74	12.71 ***
Tanzania	0.65	0.27	0.35	0.73	27.27 ***
Togo	0.52	0.07	0.48	0.93	26.13 ***
Uganda	0.66	0.87	0.34	0.13	.
Zambia	0.65	0.36	0.35	0.64	8.25 **
Zimbabwe	0.56	0.86	0.44	0.14	139.03 ***

*** p-value < 0.001, ** p-value < 0.01, * p-value < 0.05, . p-value not computed

Table 4: Proportions of violent and nonviolent events per country