Supplementary Information: Regions at Risk

Sebastian Schutte

1 Replication

You can download all replication data and code from the PSRM dataverse. Due to the large number of separate files necessary for the replication, it might be more convenient to download the zip archive at `http://sebastianschutte.net/?page_id=63`. If you want to apply the presented method to new data, all necessary functions are implemented in the "oracle" package in R which is also available on my website and on CRAN.

2 Summary statistics for the main independent variables

Table 2 on page 6 shows descriptive statistics for the independent variables used to fit the Point Process Models. Please note that this overview refers to the observations from the georeferenced covariate information available for entire countries and not just the sample of empirical points.

3 β-estimates for the PPMs

	Country	Intercept	Pop.	Cap. dist.	Access.	Brd. Dist.	Wealth	Veg.
1	Cote d'Ivoire	5.95	0.00	0.00	-0.01	-0.00	-3.33	-0.04
2	Liberia	3.94	0.00	0.01	-0.01	-0.00	0.60	-0.02
3	Guinea-Bissau	6.12	0.00	0.03	-0.02	-0.07	-8.13	-0.03
4	Sierra Leone	3.07	0.00	0.02	-0.00	-0.01	12.61	-0.02
5	Algeria	3.22	0.00	-0.01	-0.00	0.01	0.11	0.02
6	Burundi	2.77	0.00	-0.01	-0.00	-2.38	-2.86	0.03
7	Rwanda	0.88	0.00	0.00	-0.00	-1.81	-1.24	0.01
8	Congo	3.52	0.00	-0.01	-0.01	-0.15	1.53	-0.00
9	DRC	-0.60	0.00	0.00	-0.01	-0.11	3.33	-0.00
10	Chad	-2.52	0.00	0.00	-0.00	0.06	3.18	-0.01

Table 2: Overview of the β-estimates for the Point Process models. Please note that estimates are rounded to two decimals. As the models are biased due to the omission of point-to-point interactions (see main text and section 5), the standard errors cannot be substantively interpreted. The predictive capabilities of the variables are discussed at length in the main text.

4 Summary of the prediction setup

Figure 1 on page 7 clarifies the empirical setup from data preparation to simulated prediction.

5 Discussing more complex models

As described in the main text, inhomogeneous spatial Poisson models rely on one crucial assumption about the data generating process that is unlikely to hold in this application: independence of events across tiles in the spatial window (i.e. the country polygon). Of course, plenty of both qualitative and quantitative evidence suggests that conflict events provoke additional conflict events both locally and maybe even across regions within the same country. As a result, the fitted Poisson models tend to be biased as all conflict events are modeled as a function of purely local geographic conditions.

In order to incorporate point-to-point interactions, I fitted a series of "nested" models to the empirical data. In these models, additional parameters were estimated for the probability of one event generating a nearby "offspring" event and for the distance to this additional event. In multiple simulation runs, I found that this approach is not practical for two reasons. First, neither theory nor data can tell a convincing story about the maximal distance between trigger and response event. Using the Berman-Turner algorithm and the given data sources, estimates vary widely between countries. More importantly, the numerical simulations of event densities were very unstable in direct comparison to the Poisson models. This is only plausible given the fact that they rely on nested stochastic processes: in a first step, the location for the trigger event is selected and then the location for a response event is chosen. Clearly, two runs of the same simulation can easily result in two very different point patterns.

The approximation of one average density surface would have required many more than the 100 simulation iterations that I ran for each country. An additional problem is that the density estimation for the error scores rests on the assumption that the underlying point pattern has been generated through a Cox Process such as the one used in the simulation setup. Diggle (1985,147) observes that the proposed method does not easily generalize to point patterns that include point-to-point interactions as "clustering of points might be mathematically indistinguishable from variation in local intensity". In sum, sticking to Poisson processes seemed advisable in the light of mathematical and computational feasibility.

6 Results from a model averaging setup

In order to assess the relative performance of each predictor, a model-averaging setup was implemented. I fitted 63 models per country, i.e. every possible combination of the six explanatory variables excluding the specification without any predictors. For each specification, the in-sample normalized absolute error was approximated through simulation as described in the main text. The table below shows cumulative prediction errors per country for those models that included the specified predictors. Not surprisingly, geographic information on population counts produces the best predictions, i.e. the lowest error scores. The second most-important predictor in this setup is accessibility measured in the infrastructure-adjusted travel time to the next major city in the year 2000 (see main text). After that, distance to the capital city is the third important predictor. One interesting implication is that these predictors are rather static. Predictions beyond the empirical sample and into the future seem to be possible give that these important predictors are not subject to rapid change.

Var.	Av. Error	Iv. Coast	Liberia	Guinea-B.	S. Leone	Algeria	Burundi	Rwanda	Congo	DRC	Chad
Population	0.1048	0.2423	0.1090	0.0471	0.1266	0.0217	0.0995	0.1390	0.0886	0.1233	0.0507
Access.	0.1168	0.2417	0.0606	0.0506	0.1816	0.0153	0.1522	0.2227	0.0458	0.1140	0.0837
Cap. dist.	0.1188	0.2603	0.0984	0.0947	0.1810	0.0075	0.1060	0.2299	0.0426	0.0855	0.0824
Wealth	0.1259	0.2610	0.1094	0.0819	0.1822	0.0147	0.1421	0.2379	0.0699	0.0986	0.0616
Vegetation	0.1297	0.2362	0.1053	0.0789	0.1802	0.0134	0.1594	0.2309	0.0800	0.1260	0.0866
Border dist.	0.1340	0.2642	0.1100	0.0892	0.1850	0.0347	0.1448	0.2354	0.0792	0.1128	0.0851

Table 3: This table shows results from a full model-averaging setup: 63 possible combinations of the 6 covariates were estimated, while the specification without predictors was excluded. For each combination, 100 simulation runs were used to calculate the normalized MAE of the prediction surfaces as described in the main text. Average error scores are reported for each variable, i.e. those model specifications that included the corresponding variable. Errors scores are reported for individual countries, as well as the entire sample.

7 Extrapolated conflict zones

As the presented setup lends itself to predictions beyond the empirical sample, the model with the lowest out-of-sample error (model 7) was used to predict conflict zones for all of Africa as well as the Greater Middle East. While this prediction underscores the merit of the approach to inform decision making, it is critically important to be very explicit about the limitations of these predictions. First, the scope of the empirical data is limited to Africa while the predictions include the Middle East. The spatial dynamics of ongoing conflicts in these regions might differ considerably. Moreover, the model is tailored to predicting conflict zones in popular peripheral insurgencies. As explained in the main text, not all civil conflicts fall into this category. Bearing these limitations in mind, I believe that these geographic predictions of high intensity conflict zones are better than uninformed guesses. Please also note that South Sudan is not yet incorporated as an independent country. Figure 2 on page 8 shows the predictions and a corresponding KML file for visualization in Google Earth can be downloaded at `http://sebastianschutte.net/?p=325`.

Another approach shown in Figure 3 on page 9 was chosen in a second step: instead of sticking to the country boundaries and generating one prediction per sample, a single prediction was generated for the entire African continent. This approach also highlights a key advantage of the framework. Instead of relying of fixed and given polities, predictions can be generated for arbitrary regions at any scale. Interestingly, some conflict zones notoriously unstable regions are correctly reflected in this continental prediction. For example, eastern DRC, as well as the southern Uganda, Burundi and Rwanda show higher concentrations of simulated conflict events. So do the densely populated coastal areas of Northern Africa. A rigid assessment of the predictive capabilities of this approach would be necessary, but this is beyond the scope of this paper.

Descriptive	Pop.	Cap. dist.	Veg.	Access.	Wealth	Bord. dist.
Cote d'Ivoire						
Mean	31640.95	156.11	85.04	308.77	0.24	22.96
S.D.	82826.11	93.83	27.40	236.45	0.34	93.22
Min.	0.00	1.00	0.00	1.00	0.00	0.00
Max.	1657261.00	1086.00	100.00	2051.00	1.56	1086.00
Liberia						
Mean	24855.73	164.07	69.91	263.55	0.05	2.49
S.D.	31062.33	84.15	41.39	161.86	0.08	25.97
Min.	0.00	37.13	0.00	3.00	0.00	0.00
Max.	239067.50	398.00	100.00	1274.00	0.43	398.00
Guinea-Bissau						
Mean	24535.13	139.43	69.93	309.53	0.07	54.60
S.D.	35458.10	117.06	41.45	130.10	0.07	141.09
Min.	0.00	32.78	0.00	63.00	0.00	0.00
Max.	286924.19	685.00	100.00	767.00	0.21	685.00
Sierra Leone						
Mean	38404.79	108.46	83.48	233.77	0.06	6.64
S.D.	54539.33	63.21	31.06	137.40	0.04	50.20
Min.	0.00	2.31	0.00	14.00	0.00	0.00
Max.	585522.19	678.00	100.00	1048.00	0.13	678.00
Algeria						
Mean	11510.83	589.55	9.42	1342.02	0.37	32.21
S.D.	48091.79	215.76	16.54	1092.32	1.12	149.86
Min.	0.00	6.00	0.00	0.00	0.00	0.00
Max.	1892694.00	3999.00	100.00	6628.00	10.91	3999.00
Burundi						
Mean	151711.27	239.81	86.40	229.27	0.15	1.36
S.D.	111906.55	294.76	25.98	176.13	0.22	14.18
Min.	0.00	2.51	0.00	1.00	0.02	0.00
Max.	549082.38	842.28	100.00	916.00	0.64	240.00
Rwanda						
Mean	184018.70	258.02	88.89	241.17	0.20	2.25
S.D.	116115.86	327.73	21.55	164.34	0.23	26.22
Min.	26226.57	1.01	0.00	7.00	0.05	0.00
Max.	538095.31	874.94	100.00	928.00	0.64	453.00
Congo						
Mean	12833.85	445.28	92.28	962.30	0.07	277.34
S.D.	133664.72	503.08	12.36	697.99	0.16	572.00
Min.	0.00	8.24	0.00	6.00	0.00	0.00
Max.	4244614.00	3987.00	100.00	5296.00	1.19	3987.00
Congo, DRC						
Mean	16689.10	484.84	92.38	633.69	0.04	54.06
S.D.	68651.25	336.23	18.10	535.70	0.10	287.46
Min.	0.00	0.76	0.00	1.00	0.00	0.00
Max.	4244614.00	3987.00	100.00	5296.00	0.87	3987.00
Chad, DRC						
Mean	5172.50	487.99	41.68	1244.53	0.02	105.72
S.D.	10701.51	343.12	40.65	1090.61	0.04	386.16
Min.	0.00	1.62	0.00	3.00	0.00	0.00
Max.	99879.54	2990.00	100.00	6469.00	0.53	2990.00

Table 1: Descriptive statistics per country for all covariates.

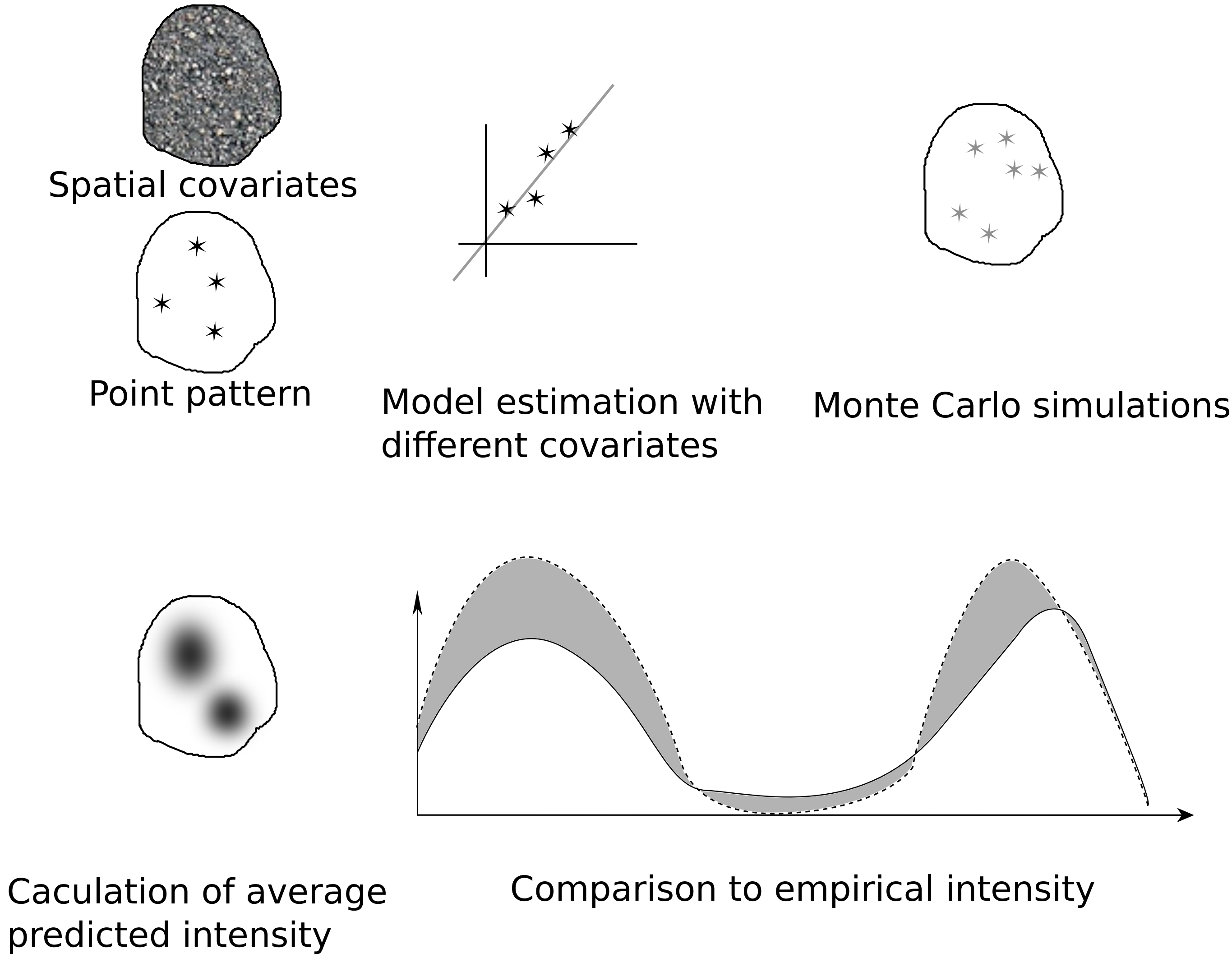

Figure 1: Overview of the empirical setup. In a first step, spatial covariates and the empirical points are loaded. After that, a series of PPMs are estimated for different convariates. In Monte Carlo simulations, average predicted intensities for the models are established. These intensities are then compared to the empirical record. As described in sections 6.1 and 6.2, comparisons are conducted in-sample and out-of-sample.

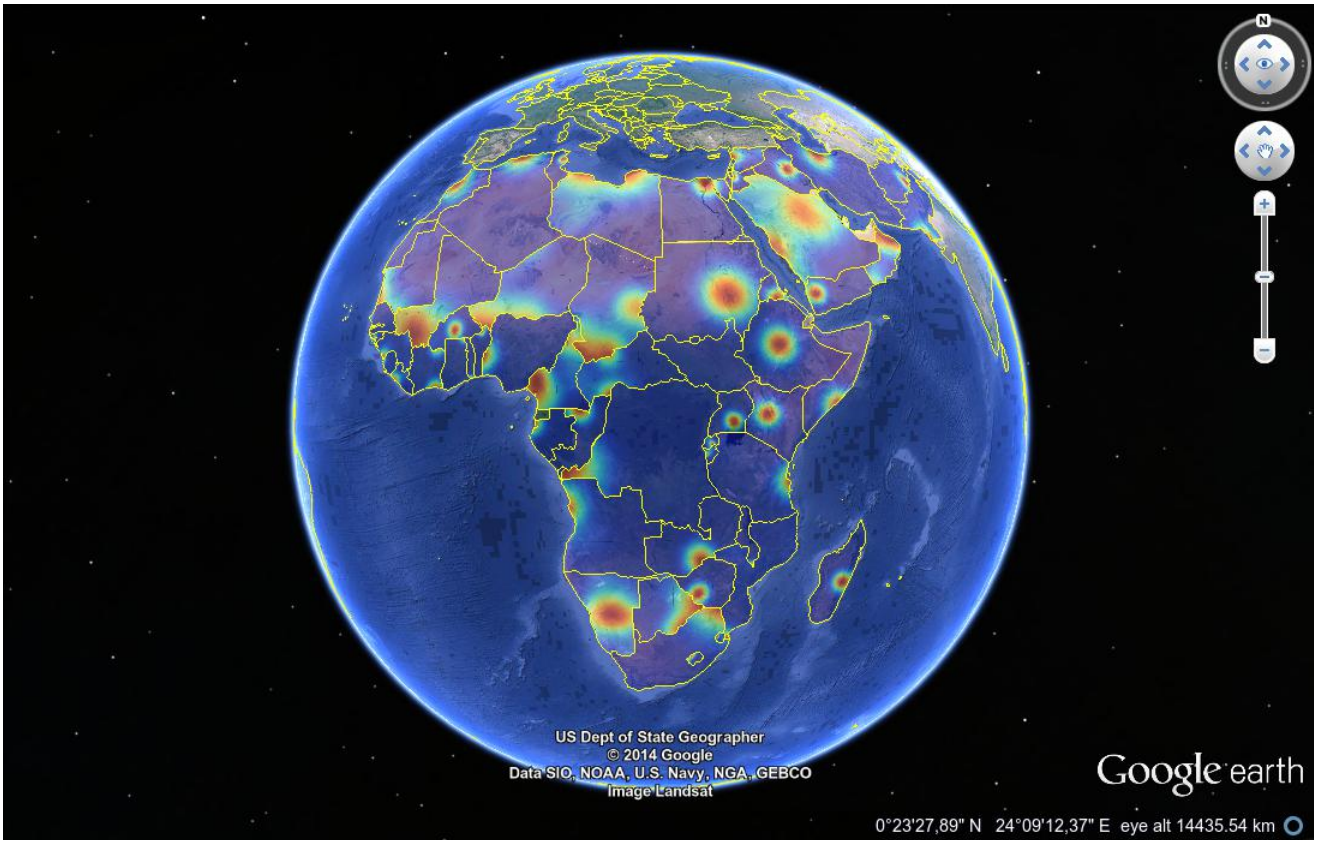

Figure 2: Extrapolated conflict zones for the entire African continent, as well as the Greater Middle East. A corresponding KML file for visualization in Google Earth can be found at http://sebastianschutte.net/?p=325.

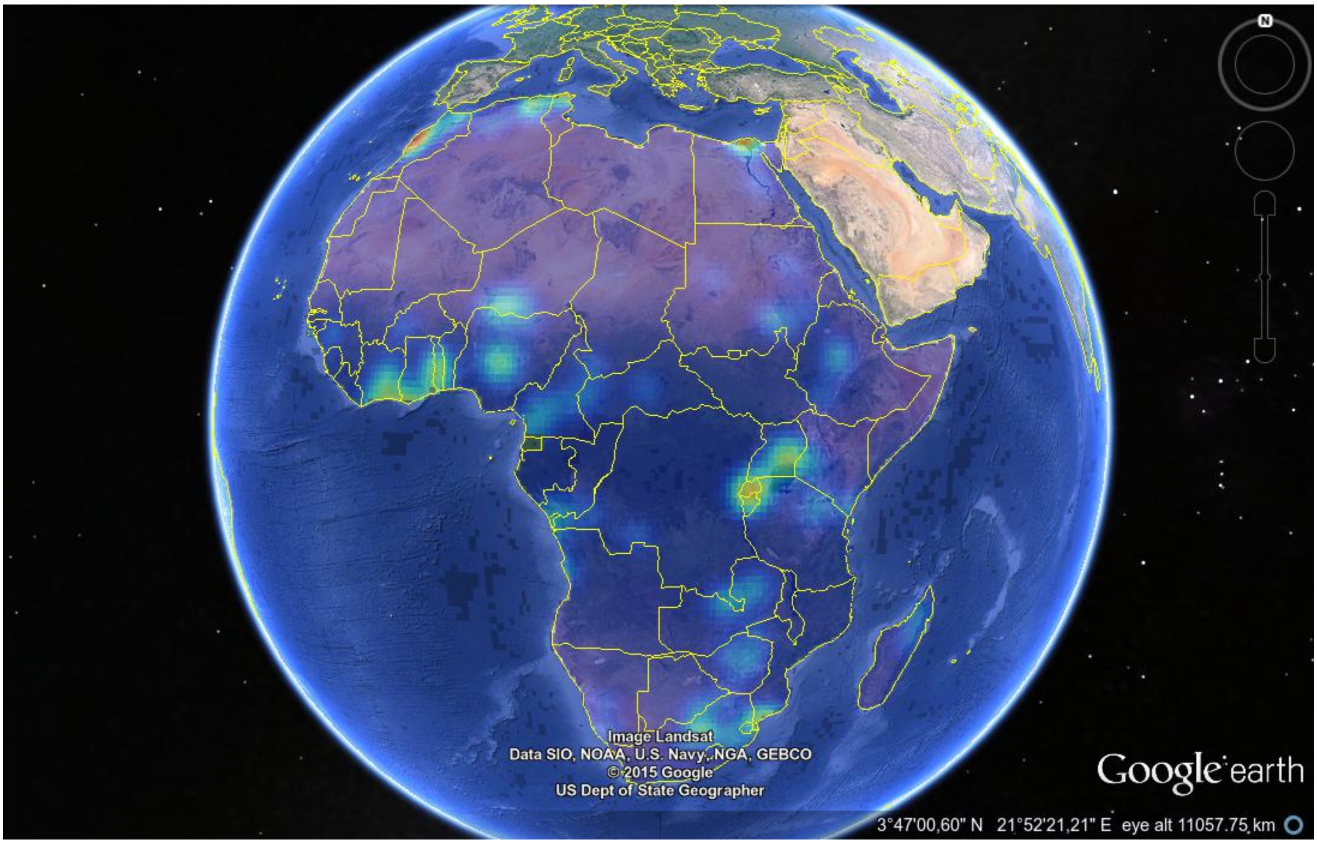

Figure 3: Extrapolated conflict zones for the entire African continent treated as a single spatial area.