Monitoring and Predicting African Countries Stability Index According to Events Data

github: `https://github.com/gzemo/africa-stability-spatial-analysis`

***Abstract*—The present analysis tests a spatially oriented hypothesis over event data collected in a 3 years period within a subset of countries belonging to the African continent. Spatial autoregressive model assessment had been computed with the purpose of evaluating whether significant and persistent relationships between event predictors and countries' stability index can be found throughout the period under examination.**

***Index Terms*—Africa, Stability Index, Spatial Durbin Model, GDELT project.**

I. Introduction

In the aftermath of the coercitive European colonialism, African countries had been generally experienced authoritarian political seasons characterised by unstable regimes that resulted in an overall generally poor administrations and significantly hindered crucial opportunities to further improve welfare and civil rights development throughout the continent [2]. This may explain the context that portray the general under-developed condition that had been progressively worsened by additionally considering intrinsic ethnic differences among populations whose relationships severely deteriorate due to local clans disputes that frequently lead to clashes and mass killings, jeopardizing any chance of future recovery, welfare and quality of life due to the direct incidence of mental discomforts like mass Post-Traumatic Stress Disorders [7].

The impact of consistent extreme climate events is also thought that will be providing a progressively negative outcome on political stability indicators yielding welfare and development reduction that will further promote unemployment and poverty [11]. Significant climate change effects are moreover exacerbating countries development by fostering severe epidemics outbreaks, like Dengue and Malaria, with an overall increasing incidence of diarrhea-related illnesses; traditional farming and water availability are also found to be seriously affected, threatening food security and increasing the dependency from foreign imports and humanitarian aids [5].

Despite the overall daunting premise, foreign investors had never quit trading goods and established commercial partnerships with local national administrations, proving that the future geopolitical scenario will be likely to highlight a new season of influences from international actors such as the Chinese government due to the extensive cooperation with Egypt and Ethiopia [8] as well as the European Union in its recent *Sustainable Development* plan[1].

[1]Among pace and security, investments are though of supporting the continent needs in light of energy and transportation sustainable transition, including digital transformation and further supporting health and education (https://www.consilium.europa.eu/en/meetings/international-summit/2022/02/17-18/).

The task of predicting whether countries may either improve or worsen their political stability is therefore showing increasingly interest: this can also be explained in light of the recent growing of specific event data sources which were found to be mostly suited to develop models for forecasting purposes[3]. Therefore, regressing future stability index may come useful for a wide plethora of humanitarian-oriented missions aiming to tackle future crisis in an promptly informed manner by gathering and allocating resources in advance.

With this being said, the present work aims at mapping the relation between historical event data and Africa's countries stability index by testing whether resulting effects may be better explained considering spatially modelled components.

A. Motivation and hypothesis

One may ask whether there could be some mutual influence while estimating a given country's stability index with respect to the behaviour of other countries in the closest proximity. Strictly speaking, the higher the presence of high gross product with stable economy states in the surrounding, the more likely any other country is to be prone to behave coherently. This may also be true on the contrary given that having less chance to develop solid relationships with neighbours may in principle hinder any development attempt. Moreover, dramatic events such as clashes between armed militias or systematic terrorism activity may worsen the country development opportunity and a spillover effect can eventually be detected considering the influence those facts might have, for instance, by abruptly reducing goods exchange or due to exodus of civilian refugees applying for asylum.

Assuming any spatial relationship, one may ask whether regression models that include spatial predictors, with the purpose of exploiting spatial dependencies, would be more able to predict the dependent variable and thus forecast stability index with respect to ordinary linear models.

Validating a feasible regression model may therefore be also useful in light of future studies' needs for exploiting limited amount of event data to predict the stability index outcome in a shorter time period by ideally reducing it from yearly to quarterly predictions. The current analysis aims to shed some light with respect to the former questions: exploratory and statistical spatial tests had been computed to address the hypothesis of interest.

II. Methods

A. Data source description

The GDELT project is a massive data-coding facility that exhaustively codifies worldwide events into pre-defined classes.

It works by translating and processing real time information produced by a wide range of sources with the purpose of defining the category of the event currently happened and which pair of principal *Actors* is involved according to the *Conflict and Mediation Event Observations* (CAMEO) code. [10]. Each event record is associated to its geo-referenced position (expressed by `{lat, lon}`) and it is described by a set of `EventCodes` that ubiquitously classify the current fact into a comprehensive list of event classes, allowing to easily collect and merge those that may share similar features or that were found to generate similar consequences.

Any record is automatically associated to a *tone* score that suggests an insight on the average tone used by the first news published on that specific event: it may take either a negative or positive value that ranges between -100 to 100 according to the qualitative assessment of the emotional magnitude involved in the said fact. Each event classified into a given category is associated to a *score* which defines the degree to which those actors involved were more likely to either cooperate or compete, ranging from a maximum positive score (+10) to a minimum (-10) respectively (Goldstein, [6]). This value aims at providing an ordered set of categories of international affairs sorted by an absolute relevance measure which may in principle define either cooperation or crisis between actors in either "verbal" or "material" manner.

The *Worldwide Governance Indicators* are unique indexes computed by the World Bank Group[2] that seek to quantify a variety of countries governance key factors through time in order to express scores that reflect Accountability, Corruption, Government Effectiveness and Stability. To this extend, the *Political Stability and Absence of Violence/Terrorism* index provides an insight on the likelihood of a country to experience political instability that may further evolve into politically oriented clashes and terrorism.

Finally, the corresponding Africa geometry had been downloaded from the *ArcGIS Hub* in a ESRI Shapefile in EPSG:4326 projection (Geodetic CRS: WGS 84).

B. Data processing

The current work delves into the amount of geo-referenced events data collected from 2020 to 2022 (included) and filtered by retaining those records whose coordinates were found to be within the continents' geometry by iteratively performing an inner join operation on records files over the aforementioned time-span. Coherently only those countries, whose Stability Index was found to be estimated for all years of interest, had been retained in the final `GeoDataFrame`.

At each record event that was found eligible to be included within a given countries' geometry, the corresponding country reference name was associated in order to identify the location of the fact and to bind it within an uniquely defined state's border regardless the original GDELT *Actor* assignment system[3].

C. Neighbours estimation

The *Spatial Weight Matrix* (W) had been built and row-standardise by considering all neighbouring countries in a *Contiguity-based* manner in order to model any possible spatial dependency according to the real amount of borders shared in common by adjacent regions: this implies that groups of islands such as the Republic of Madagascar or the Republic of Mauritius are excluded from the analysis. The choice's rationale relies on the assumption that facts that may enhance or threaten any country stability are more likely to be related to crossing-border events like trading goods, infrastructure construction or refugees/immigration flows.

Two approaches were tested in order to define the neighbouring regions weights ($w_{i,j} : i \neq j$): *a)* by *statically* assigning the same weighting factor allowing each state to equally contribute and *b)* by *dynamically* considering the absolute value of the yearly averaged *Tone* value of all news for which the $i-th$ and $j-th$ countries were found to be detected and codified as Actors involved in such facts resulting in an index, normalised by the magnitude of the influence, that varies over years and aims at modelling countries relationships.

D. Predictors definition and estimation

The complete set of records had been filtered by the corresponding date in which each fact was originally registered in order to extract the yearly predictors described as follows:

`EventRootX`: according to the CAMEO code, events can be classified into categories that reflect a given relevance and are similar with respect to the consequence of the actions involved. An attempt to directly model the way in which the country's stability index could be influenced by the magnitude of occurring happenings is to count the amount of recorded geo-referenced facts belonging to a certain set of event classes and use them directly as predictors. In other terms, the actual number of some kind of event that happened within a specific administrative border and gathered in a certain period of time is supposed to play some role while predicting the overall country stability index. Specifically, all events belonging to the following categories were collected:

- *"Diplomatic cooperation engagement"* (CAMEO code `5`) defines a class of events that includes condition in which governments or political parties express the intention or directly behaves in order to sign agreements with other local or international institutions in order to support them.
- *"Material cooperation engagement"* (CAMEO code `6`), as the former it provides a pool of events for which an existing material exchange between governments is occurring in either economical, military or judicial manner.

[2] A global partnership that is committed in the improvement of third-country economies, seeking to fight against severe hunger and poverty (https://www.worldbank.org/en/who-we-are).

[3] The automated data-coding algorithm is able to identify Government as well as Non-Government associations, Religious and Ethnic groups, Regular military and paramilitary groups involved in the codified events. However, some records detect actions that were made or experienced by other players that may not be directly bounded by a given Government administration.

- *"Provide aid"* (CAMEO code `7`) includes events like *"Providing Humanitarian Aid"* or *"Providing military protection or peacekeeping"* underling an actual intent that has taken place.
- *"Demand"* (CAMEO code `10`) presents a set of events in which an actor requires an action to be taken immediately or after it had already been encouraged in the past an the corresponding outcome expectations had not been satisfied.
- *"Reject"* (CAMEO code `12`) depicts any sort of events in which either verbal or material cooperation had been rejected like *"Reject judicial cooperation"* or *"Reject request for humanitarian aid"*.
- *"Threaten"* (CAMEO code `13`) provides a class of facts that take into account verbal acts aiming at threatening to impose and actively behave such as *"Threaten political dissent"* or to ban, reject, disapprove like *"Threaten to ban political parties or politicians"*.
- *"Protest"* (CAMEO code `14`) defines all riots-related events with either pacific outcomes like dissent or strikes or violent consequences.
- *"Fight"* (CAMEO code `19`) which provides an idea of any conventional violence event including military engagement.

`AvgTone`: refers to the yearly average news tone ($tone$) estimated on each set of country's neighbours which namely provides a measure of the relevance of all events that had happened in the surrounding proximity for each given state regardless their actual event code. Specifically, within a given year $year \in \{2020, 2021, 2022\}$ and for each country $C_i : i = 1, .., N_{countries}$, the complete set of events $E_i^{[year]}$ that were found to be located within the administrative borders of all $N_j : j = 1, .., |neigh(C_i)|$ (the country's neighbours[4]) were retained in order to extract the associated average $tone$.

$$avgtone_i^{[year]} = \frac{1}{|E_i^{[year]}|} \sum_{e=1}^{|E_i^{[year]}|} tone_e$$

This predictor gives some clues on the average tone of all facts that, although are not supposed to be directly associated to the corresponding country of interest (C_i), are defining either a positive or negative trend of what had occurred in its proximity. Along with the predictor that will be discussed as follows, this variable formalizes a measure that is supposed to give an insight on the surrounding countries stability.

`GoldsteinScore`: similarly to the estimation of the former averaged tone, the $i-th$ country yearly averaged Goldstein Score is computed by retaining the score (GS) of all events happened in the surrounding country's neighbourhood yielding $avgGS_i^{[year]}$.

[4]Ideally, given that borders may change over time, the corresponding geometry of all N_j neighbours of a given C_i country should have been computed by taking also into account any yearly country's border change in order to filter those events that may have been localized within the $N_j^{[year]}$ neighbours borders to address any possible variation in the set of adjacent neighbours of a $i-th$ state. The present analysis neglected this aspect given that the purpose of the present work delves into fairly recent events for a limited amount of time and significant changes in states' borders is not expected to drastically change the neighbours detection for a given country.

E. Exploratory analysis

The frequency of a certain set of event categories had been extracted and plotted over time showing the top 10 countries ordered by the highest amount of overall events occurred within its boundaries over the period under examination. Global Moran index had been assessed to verify any form of Stability Index spatial dependency. Accordingly, Local Moran test and Local Indicator of Spatial Associations [1] had been computed in order to understand how spatial data are organized within the continents by investigating areas that can be clustered together with respect to similar dependent variable's values and by addressing whether clusters that present unusual concentration of values are significantly-unlike to be displayed while considering random spatial data distribution [9].

F. Statistical analysis

For each year, the corresponding subset of records had been filtered and tested individually. The dependent variable (Stability index) had been firstly examined according to the Moran Index test in order to verify whether a significant spatial relation can be found considering the current set of neighbouring regions modelled according to either the *static* or *dynamic* approach. Following, according to the Helhorst procedure [4], a Lagrange Multiplier (LM) test had been estimated in order to verify which condition among the Lag or Error inducing global or local spillover effects respectively could have better described the existing spatial relationship against the Ordinary Least Square (OLS) model which does not take into account any spatial relationship. According to the LM results, a *Spatial Durbin Model* (SDM), formalised as follows:

$$y = \rho W y + X\beta + W X\theta + \epsilon$$

had been fitted in order to be tested by Likelihood Ratio test (LRT) to evaluate whether spillover parameters θ and ρ are significantly different from zero thus helping to further accept SDM in favor of the simplier *Spatial Autoregressive Model* (SAR) or *Spatial Error Model* (SEM). Impact assessment testing the Direct, Indirect and Total impacts had finally been performed to explain coefficients found to be significant in light of the spatial dependency.

III. RESULT

Moran scatter plots show positive relationships considering the Stability index over years (*see Figure* 1). Although the *static* as well as *dynamic* spatial matrix weighting approaches both proved to yield statistically significant Moran Index tests, the former was consistently found to achieve more robust probability (not shown). Therefore the technique first discussed was adopted for the following analyses.

The frequencies of average daily events count is shown as in *Figures* 2, 3 displaying for most of the 10 top countries rather constant fluctuations over time except for some countries like Sudan showing some spikes periods in which after a significant change in the amount of daily events it approximately returns to the baseline amount of facts in both riot- and violence-related timeseries.

Fig. 1. Moran scatter plots considering the Stability index over years as dependent variable.

The spatial pattern of the mean of the daily average amount of either riots or fight events is displayed in *Figures* 4, 5: again countries such as Nigeria and South Africa seem to pop out in both categories displaying significantly higher values with respect to those states in the closest proximity. Interestingly a coherently organized cluster showing similar data over time can be drawn considering Egypt, Libya, Sudan, Ethiopia and Kenya.

Local correlation plots and test estimated results by filtering data from 2020 to 2022 are shown in *Figures* 6, 7, 8 in which the distribution of stability indexes is compared with the results from those regions that were proved to present statistically significant different spatial values with respect to random rearrangement. Two significant clusters can be detected over times that include countries that are either mostly from central Saharian and sub-Saharian regions or from the southern portion of the continent (*see Table* I).

These clusters depict different circumstances of spillover effects in which regions presenting lower stability indexes are more likely to be surrounded by states that share similar scores and vice-versa. Indeed, the former cluster is found to include regions belonging to the bottom left Moran scatter plot region (LL) and for which a generally lower stability index is detected whereas the latter (HH) provides countries whose stability measures are coherently higher.

Interestingly, the Local Moran index clustering result seems to be further supported by the incidence of the average daily amount of riot related events in the first cluster (like in Egypt, Sudan and Ethiopia related records).

The complete commented spatial statistical analysis results can be found in `spatial_analysis.pdf`.

IV. Discussion

The spatial statistical analysis results highlighted that, although the subset of 2020 records for which a SAR model is shown to reach the same amount of goodness of fit of an OLS model, records filtered from 2021 and 2022 were best explained by SAR due to LRT results suggesting that the spatial lag parameter ρ can be significantly considered different from zero.

TABLE I
REGIONS BELONGING TO SIGNIFICANT SPATIAL CLUSTERS ACCORDING TO YEAR AND MORAN SCATTER PLOT POSITION.

Year	Position	Countries
2020	LL	Egypt, Sudan, South Sudan, Chad, Niger, Ethiopia, Central African Republic, Kenya
	HH	Namibia, South Africa, Botswana, Zimbabwe, Mozambique, Zambia
2021	LL	Egypt, Sudan, South Sudan, Chad, Niger, Ethiopia, Central African Republic, Kenya, Eritrea
	HH	Namibia, South Africa, Botswana
2022	LL	Egypt, Sudan, South Sudan, Chad, Niger, Ethiopia, Central African Republic, Kenya, Eritrea
	HH	Namibia, South Africa, Botswana, Zimbabwe

Results consistently proved to confirm the global over the local spatial spillover effect throughout all years according to the LM test results, indeed supporting the initial hypothesis.

Among the list of geo-referenced variables chosen to model the stability index prediction the amount of fight (*EventRoot = 19*) and material cooperation events (*EventRoot = 6*) were found to generate significant impacts over the Spatial models estimated from the 2021 and 2022 data respectively. By contrast, the former result highlighted generally a negative Direct Impact (-0.0001, $p - value < 0.001$) proving that the higher the amount of events coded as belonging to the root code 19, the more likely to have stability indexes to be affected in a negative manner within each country. The latter, proved that a positive Direct Impact (0.002, $p - value < 0.05$) is enhancing a material cooperation events dependent effect showing that the higher the occurrence of goods trading, the higher the stability index of that country on average.

Although spatial autoregressive models generally improved the OLS, no global spatial spillover effect was found due to the lack of significance level in any predictor's Indirect Impact, yielding virtually no evidence that may support the current work's hypothesis.

No spatial model predictors were found to result in statistically significant impact over all yearly subsets meaning that either a valid significant predictor had not been modelled yet or the spatial relationship can actually been explained by different factors over time. This conclusion needs further clarifications to better understand the role of the subset of predictors and their actual predictive capability.

A. Limitation and further improvements

The lack of consistent results over the time-span considered may be explained in light of the design choices taken while defining the predictors categories: selecting all events according to macro classes instead of a few more relevant and specific kind of facts may have worsen models' predictive capabilities due to an excessive amount of noise.

Predictors like *avgTone* and *GoldsteinScore* had been computed by retaining and averaging the corresponding score of all kind of events regardless their original pair of Actors involved. Further filtering improvements are needed in order to accurately select more relevant information exchange between international players relying for instance exclusively on Government Organizations or Political Parties.

No further spatial analysis had been completed with the purpose of exploring models implementing *dynamically* changing W lag matrix by weighting spatial connection via countries mutual average tone. Despite the slightly lower significance level it may have ultimately addressed time-dependant changes in the stability measure providing overall useful insights. Finally no probability correction for multiple comparison had been addressed while testing the hypothesis of interest over years meaning that weaker significant results are not expected to survive.

V. Conclusion

The attempt here presented maps countries' stability score according to an easily exploitable source of information, as offered by the GDELT project, to explore predictors capabilities and to test their consistency over time. Although not explicitly used to further evaluate spatial regression across years, a solution to model a spatial weight matrix that deals with countries relationships dynamics is presented and tested yielding promising results. Further research is needed in order to better understand whether a valid subset of event data predictors may efficiently generate significant Direct and Indirect Impacts over time: this would ultimately support evidences of significant spillover effects proving that relevant facts in the surrounding regions are playing some role influencing a given country stability score.

References

[1] Luc Anselin. "Local indicators of spatial association—LISA". In: *Geographical analysis* 27.2 (1995), pp. 93–115.

[2] Jutta Bolt et al. *African political institutions and the impact of colonialism*. Tech. rep. National Bureau of Economic Research, 2022.

[3] Charles Butcher et al. "Introducing the targeted mass killing data set for the study and forecasting of mass atrocities". In: *Journal of Conflict Resolution* 64.7-8 (2020), pp. 1524–1547.

[4] J Paul Elhorst. "Applied spatial econometrics: raising the bar". In: *Spatial economic analysis* 5.1 (2010), pp. 9–28.

[5] MS Evans and B Munslow. "Climate change, health, and conflict in Africa's arc of instability". In: *Perspectives in Public Health* 141.6 (2021), pp. 338–341. DOI: 10.1177/17579139211058299.

[6] Joshua S Goldstein. "A conflict-cooperation scale for WEIS events data". In: *Journal of Conflict Resolution* 36.2 (1992), pp. 369–385.

[7] Seggane Musisi and Eugene Kinyanda. "Long-term impact of war, civil war, and persecution in civilian populations—Conflict and post-traumatic stress in African communities". In: *Frontiers in psychiatry* 11 (2020), p. 478377.

[8] Oluwaseyi Ebenezer Olalere. "Chinese Investment in West African Infrastructural Development: Selected Case Studies". In: *Perspectives on Africa-China Infrastructural and Industrial Cooperation: Empirical Findings and Conceptual Implications*. Springer, 2023, pp. 59–72.

[9] Jeffery Sauer et al. "The importance of null hypotheses: Understanding differences in local Moran's under heteroskedasticity". In: *Geographical Analysis* 54.4 (2022), pp. 752–768.

[10] Philip A Schrodt et al. "The CAMEO (conflict and mediation event observations) actor coding framework". In: *2008 Annual Meeting of the International Studies Association*. 2008.

[11] Emrah Sofuoğlu and Ahmet Ay. "The relationship between climate change and political instability: the case of MENA countries (1985: 01–2016: 12)". In: *Environmental Science and Pollution Research* 27.12 (2020), pp. 14033–14043.

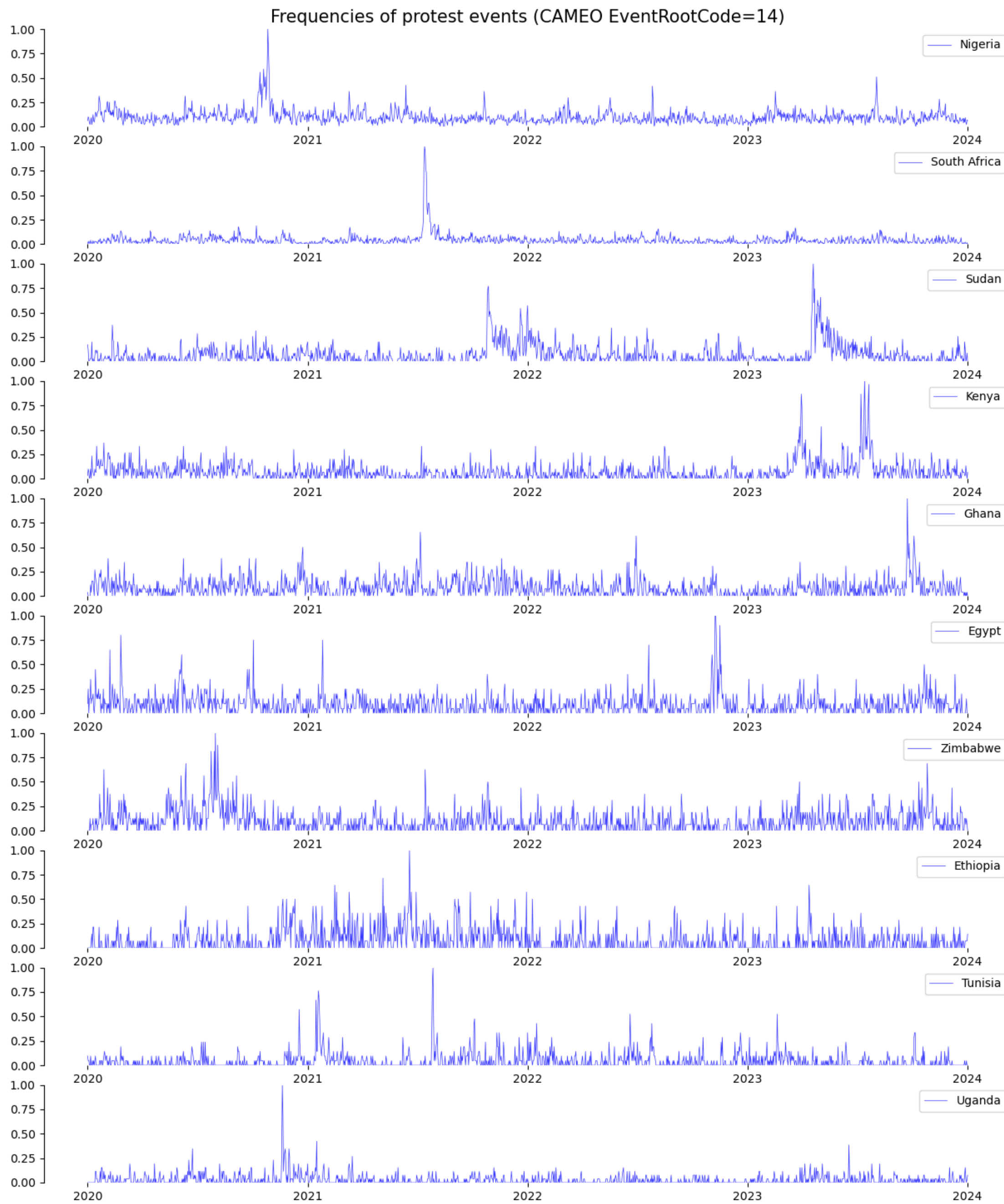

Fig. 2. Frequency of daily average riot-related events codified by CAMEO `EventRootCode = 14` over the period under examination (including 2023) of the 10 top countries that displayed the highest amount of events over time. Values are normalized in a [0,1] range by rescaling each country according to its max value.

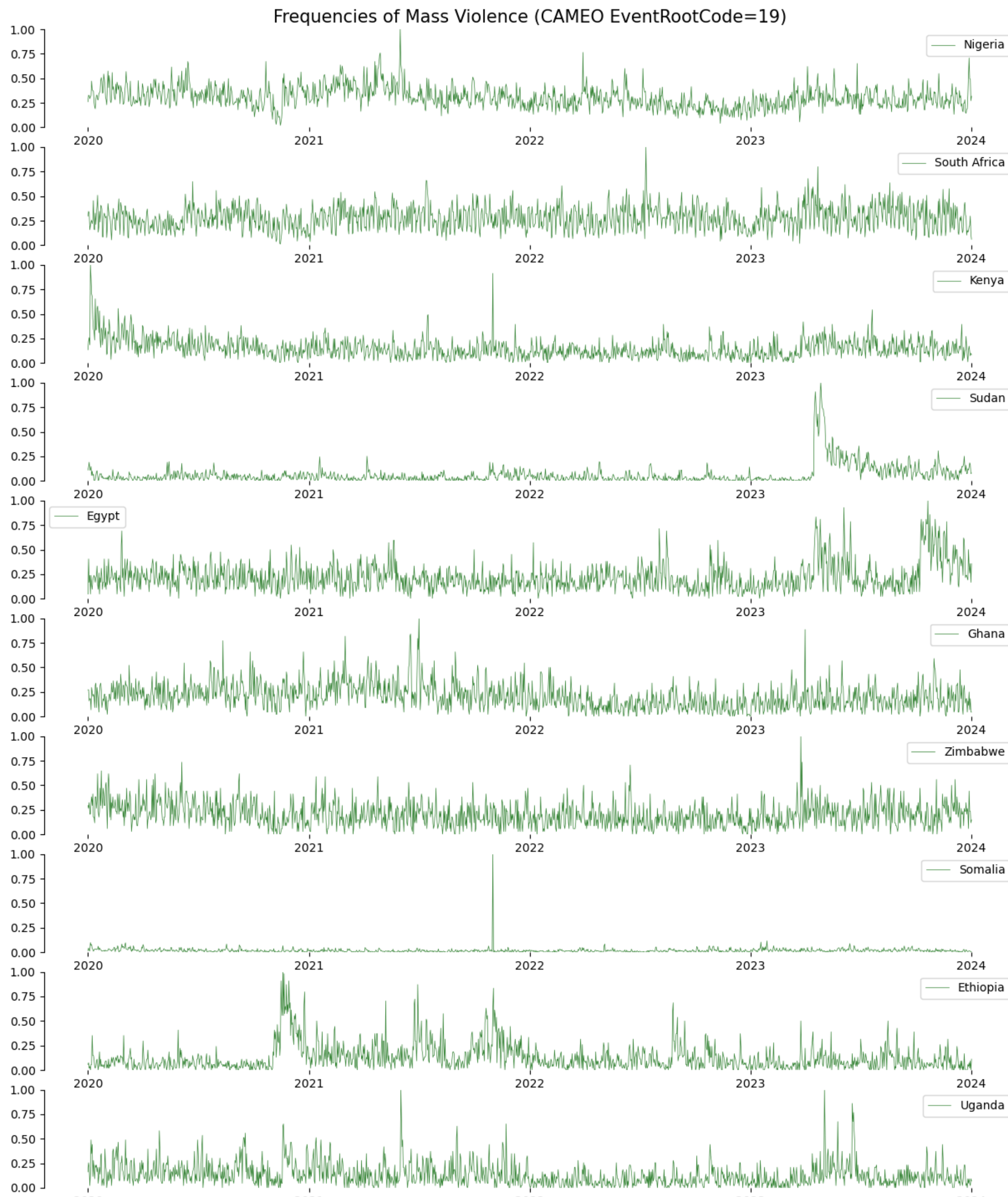

Fig. 3. Frequency of daily average Fight-related events codified by CAMEO `EventRootCode = 19` over the period under examination (including 2023) of the 10 top countries that displayed the highest amount of events over time. Values are normalized in a [0,1] range by rescaling each country according to its max value.

Fig. 4. Choroplet map depicting the spatial pattern of the mean daily average amount of riots event (CAMEO `EventRootCode = 14`) in the continent over years.

Fig. 5. Choroplet map depicting the spatial pattern of the mean daily average amount of fight event (CAMEO `EventRootCode = 19`) in the continent over years.

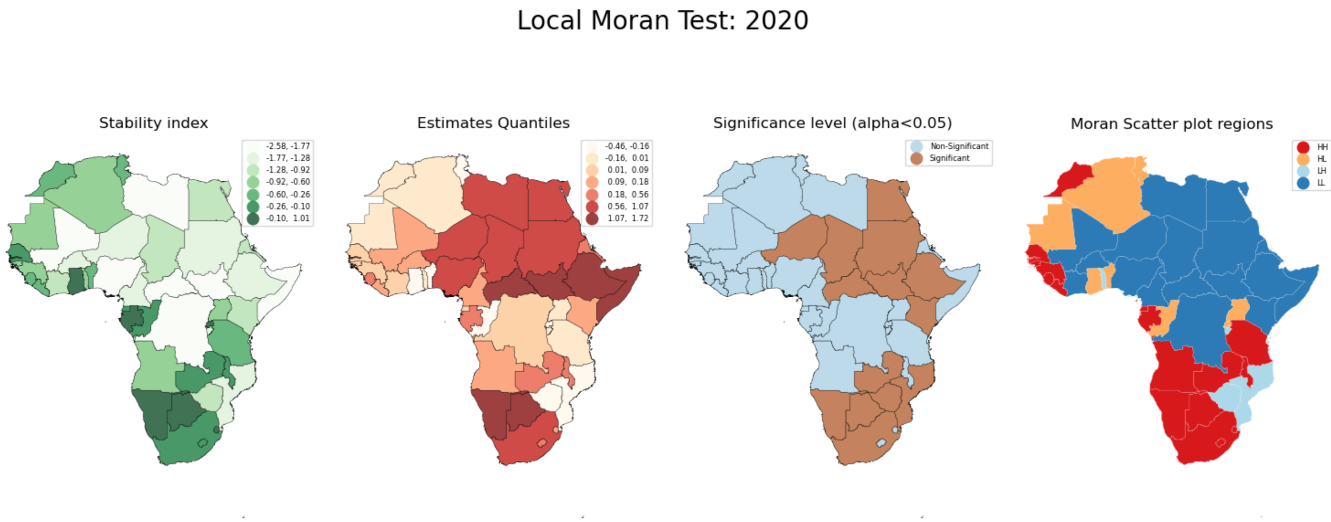

Fig. 6. Results from the Local Moran Index test performed by considering records filtered from the 2020 subset. From left to right: Stability index map and Local Moran test estimate values are displayed in by dividing the real range in quantiles and color-coded according to the scheme: the lighter the lower range of values included in that interval. Follows Significance levels and position according to the Moran Scatter plot are shown.

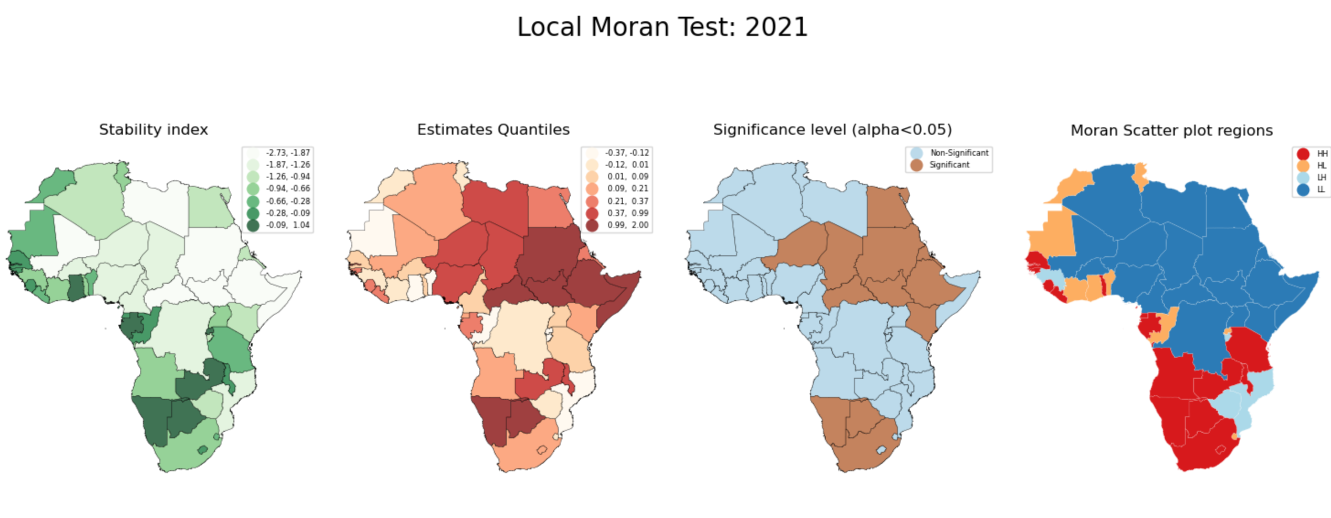

Fig. 7. Results from the Local Moran Index test performed by considering records filtered from the 2021 subset. From left to right: Stability index map and Local Moran test estimate values are displayed in by dividing the real range in quantiles and color-coded according to the scheme: the lighter the lower range of values included in that interval. Follows Significance levels and position according to the Moran Scatter plot are shown.

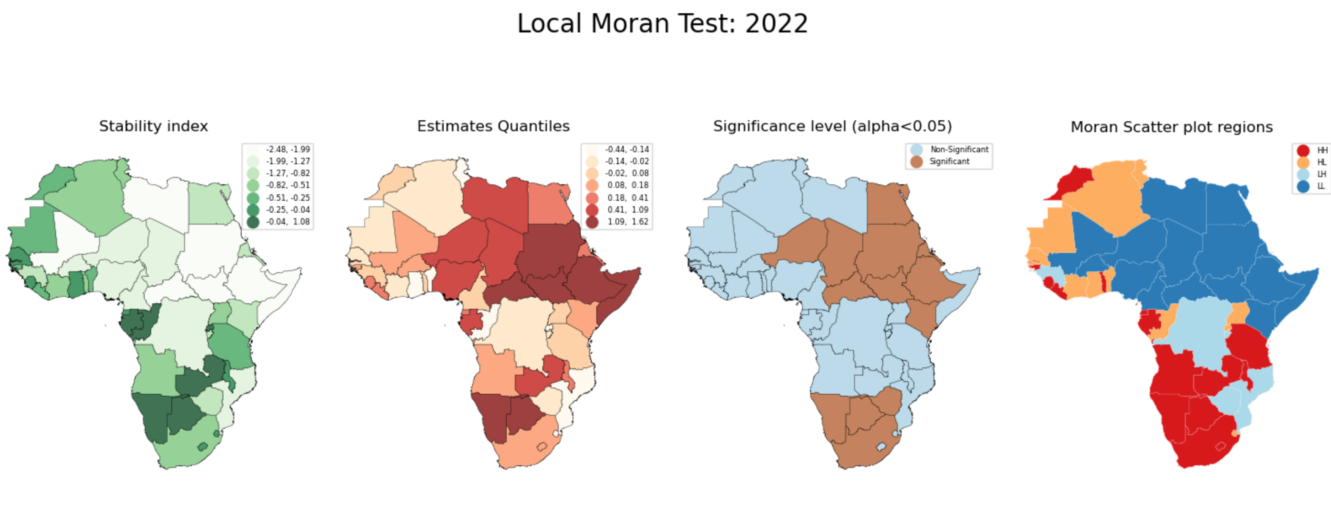

Fig. 8. Results from the Local Moran Index test performed by considering records filtered from the 2022 subset. From left to right: Stability index map and Local Moran test estimate values are displayed in by dividing the real range in quantiles and color-coded according to the scheme: the lighter the lower range of values included in that interval. Follows Significance levels and position according to the Moran Scatter plot are shown.