Computing and Informatics, Vol. 41, 2022, 463–478, doi: 10.31577/cai_2022_2_463

EARLY WARNING SIGNALS IN OPEN SOURCE INTELLIGENCE: TWO USE CASES OF THE 2019 IRAQI AND 2020 INDIAN FARMERS' PROTESTS

Étienne Voutaz, Albert Blarer

armasuisse Science and Technology S+T
Cyber-Defence Campus
Feuerwerkerstrasse 39
CH-3602 Thun, Switzerland
e-mail: {etienne.voutaz, albert.blarer}@armasuisse.ch

Abstract. Early warning signals methods have been introduced in the field of ecological sciences and widely used in other domains. However, while these methods have proven effective for deterministic dynamics governed by differential equations or smooth maps – both on synthetic and real data – their application in the social sciences is more complex. A series of protests started in Iraq on 1 October 2019 and farmers' protests in India in September 2020. We investigate in this work how these waves could have been anticipated using early warning signals for the time series of daily occurrences of protests. We use to this end metric-based indicators (autocorrelation at-lag-1, standard deviation and skewness), analyse trends using Kendall rank correlation and use bootstrapping methods to implement a statistical test exhibiting a regime shift (tipping points) in the dynamics of protests. We moreover highlight the importance of the standard deviation as an indicator.

Keywords: Early warning signals, tipping points, bifurcation, OSINT, time series analysis

Mathematics Subject Classification 2010: 62M10, 65P30

1 INTRODUCTION

Early warning signals (EWS) are methods used to study bifurcations associated with tipping points in time series. The basic idea is that when a bifurcation occurs in a dynamic system, the analysis of indicators prior to the regime shift can help predicting it. Such methods were first introduced in [16, 14] and applied in different fields like ecological science, finance or social science [10, 19, 18, 15].

The abrupt shift of an ecological system to an alternate stable state was first obtained from work on theoretical model and heavily criticized [16]. These methods have since been validated on real data. A 'fold'-bifurcation is for example depicted in Figure 1. The ecosystem cannot pass smoothly from the upper branch of the folded curve to the lower one. Instead, a catastrophic transition to the lower branch occurs.

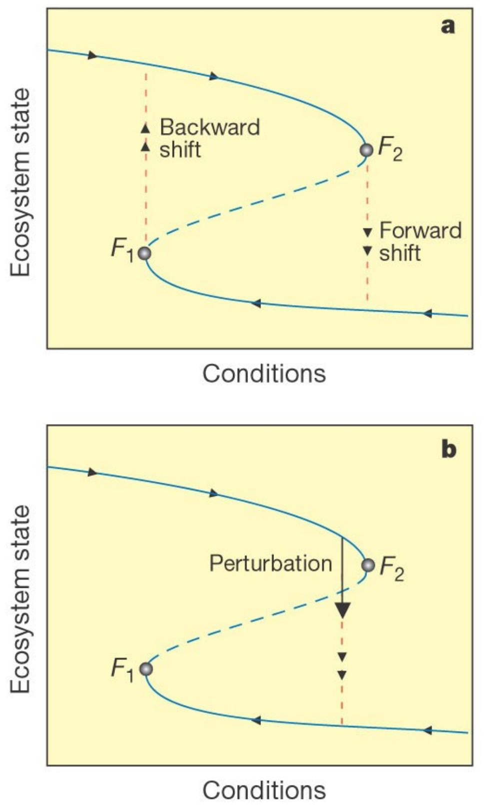


Figure 1. Shifts between alternative stable states [16]. a) If the system is close to the bifurcation point F_2, a small perturbation drives the system to the lower branch (forward shift). A backward shift occurs if the conditions are reversed to reach the other bifurcation point F_1. b) A sufficiently large perturbation close to the bifurcation point can also shift the dynamics to the other stable state.

In that case, the dynamics exhibits a low resilience (see Figure 2).

There are two main approaches to anticipate tipping points: a metric-based

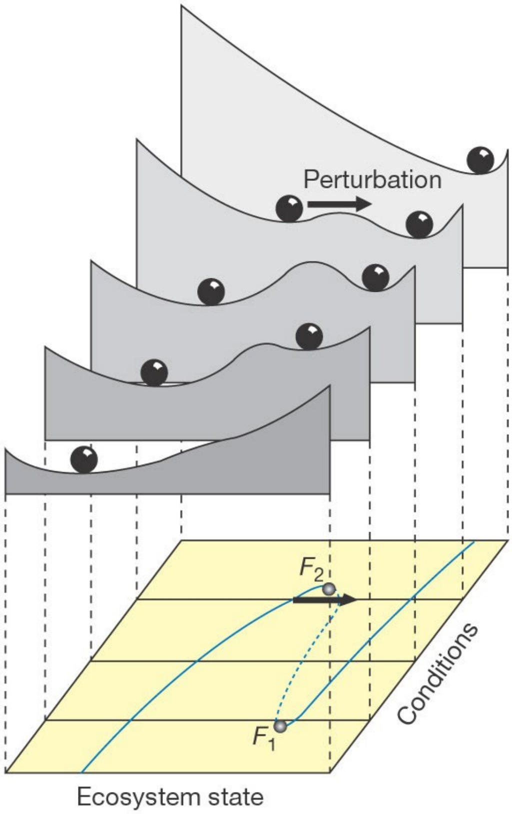


Figure 2. The system shifts from a stable state in the foreground to another stable one in the background in a non-smooth way [16]. The intermediate states show small basins of attraction and, therefore, low resilience.

and a model-based approach. Metric approaches compute indicators using standard metrics from descriptive statistics like autocorrelation, standard deviation, skewness while model-based methods first try to fit the data to a model. Autocorrelation captures the memory of time series while standard deviation and skewness indicate changes in its variability and flickering nature. Both methods aim to find bifurcation patterns in the dynamics in the vicinity of the change of regime.

Intuitively, in the vicinity of a tipping point, the system takes longer to return to equilibrium. It is like dropping a ball on a convex surface. The less convex the surface, the longer it will take for the ball to find its equilibrium, and then the equilibrium will be broken if there is no convexity any more. This slow return to equilibrium will impact indicators like autocorrelation, standard deviation or skewness: they will tend to increase.

Let us stress that the regime shifts EWS try to identify and anticipate not only represent dynamical discontinuities but 'catastrophic' changes in the underlying dynamics, e.g., a transition from one stable state to another one.

2 CONTEXT

In this work we aim to study time series coming from open source intelligence (OSINT). The data used represent the daily number of protests in Iraq and India (see Figure 3) and was downloaded from 'The Armed Conflict Location and Event Data Project' (ACLED) [6]. ACLED collects the dates, actors, locations, fatalities, and types of all reported political violence and protest events across a large part of the world.

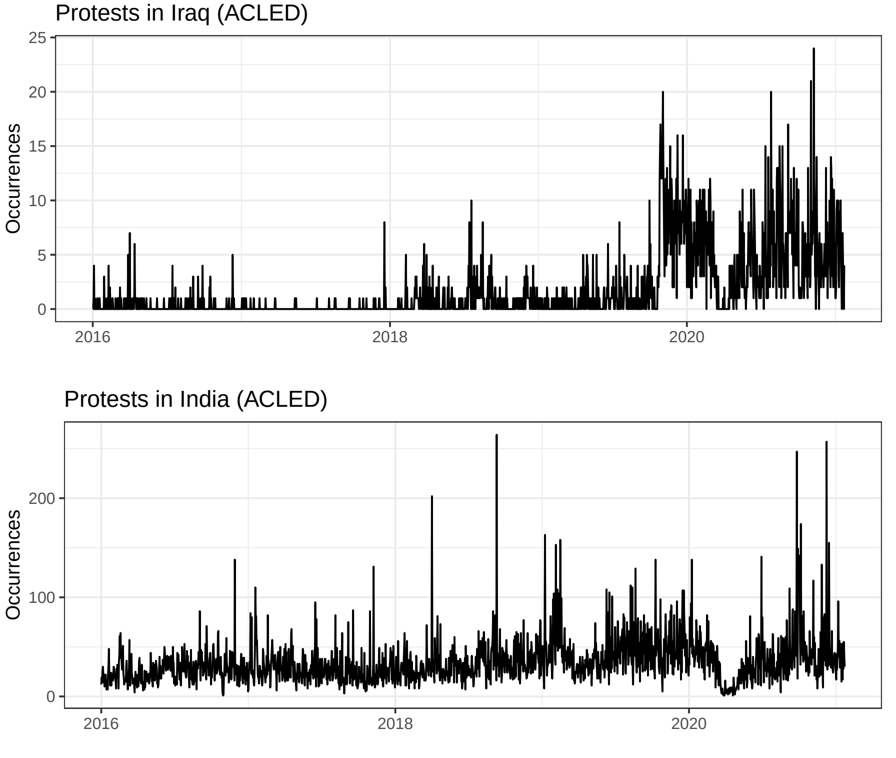


Figure 3. Daily protests in Iraq (top) and India (bottom) from 2016-01-01 to 2021-01-29. Note the difference in the intensity of protests in both datasets: the daily number of protests in Iraq does not exceed 30, with a non-negligible number of days without events; the intensity of the Indian protests is much higher (source: ACLED, retrieved on 29 January 2021).

Protests are good indicators of the social climate in a particular region of the world and trying to anticipate changes in such a time series is a natural intelligence goal.

The Iraqi dataset indicates a change around the beginning of October 2019. In fact, the 2019–2021 Iraqi protests are a series of protests that consist of demonstrations, marches, sit-ins and civil disobedience. They started on 1 October 2019, a date which was set by civil activists on social media, spreading over the central and southern provinces of Iraq, to protest corruption, unemployment and inefficient public services [2]. The protests were the largest incident of civil unrest Iraq has experienced since the 2003 invasion. The peak observed in the time series after 1 October 2019 begins on 26 October 2019.

The Indian dataset does not exhibit a clear change as in the Iraqi case, except the drop at the beginning of 2020 which is probably due to the Covid pandemic. However, an increasing flickering tendency is noticeable in the second part of 2020. A wave of ongoing farmers' protests began at the end of September 2020 against three farm acts which were passed by the Parliament of India [5].

The study of EWS in this context is a priori difficult since the underlying dynamics of time series coming from social science is not driven by physical laws. As example, the explosion in Beirut in August 2020 was followed by massive protests that could not have been anticipated due to the random nature of the event. However, the causes of the protests are not uniquely due to this exogenous, unpredictable event; root causes can be found in economical and political collapses in the Lebanon society.

3 CONTRIBUTION

Measuring the intensity of protests in a region of the world is a good measure of social happiness and social climate. However, protests are often preceded by a rise of anger, sadness or tension. It is shown in [17] using a localized set of blogs that a period of slowing down preceding tipping points can be identified.

An increase of variance and autocorrelation has been detected in social network for known events [13]. Theoretical considerations make it possible to highlight that, in the vicinity of a tipping point, the variance follows a power law. This law could be identified, but note that only a priory known events like Christmas or Halloween are considered in this work. Further important questions are moreover formulated therein:

1. How do we define when a critical transition occurs for an a priori unknown event in the data?
2. Can we link warning signs in social networks to a priori unknown critical transitions outside a social network?
3. Which models of social networks can re-produce critical transitions observed in data?

In this work we prove that EWS do exist using real data from the OSINT context described above for unknown events. We also show that the standard deviation plays an important role as an indicator.

4 DATASET

The datasets were retrieved on 29 January 2021 from the ACLED database according to the policy available at that time. Countries 'India' respectively 'Iraq' and event type 'Protests' were selected. Protests are defined as 'A public demonstration against a political entity, government institution, policy or group in which the participants are not violent'. Each event contains informations like date, description, type (e.g. protests), country, region, longitude, latitude, actors, etc. Here, we simply count how many events daily occur. Note that the localization of the events over the considered period is depicted in Figure 4.

5 BASELINE

We use as baseline a naive method aiming to anticipate the Indian farmers' protests at the end of September 2020. The idea is to use a change point detection algorithm. In statistical analysis, change point detection (CPD) tries to identify times when the probability distribution of a time series changes. Most of the time, changes in mean/variance are targeted. A large number of algorithms are used in different fields of applications. An evaluation of some of the most popular ones is given in [9].

The CPD algorithm we will use is Bayesian Online Change Detection (BOCD)[8]. Firstly because it gives good results [9] and secondly because it is online, i.e., a change point is sought since the last change point appeared.

Note that the method is naive since there is no theoretical evidence that a relation between tipping points and change points does exist.

The algorithm identifies 20 change points in the Indian time series. One of them appeared on 25 September 2020 and the preceding one on 22 May 2020. In this way the beginning of the protests is identified by the change point algorithm but not anticipated.

It is worth noting that ACLED delivers a service (Early Warning Research Hub) offering a suite of resources aimed at facilitating data-driven initiatives to anticipate and respond to emerging crises [7].

6 METHODS

The first usual step consists in stationarising the time series. So it is first log-transformed and differenced once, i.e., we transform the original time series (x_n) as $y_n = \log(x_n) - \log(x_{n-1})$. We use then rolling windows of fixed size over the time series and divide each window in two halves. The first half is used to compute indicators (autocorrelation at-lag-1, standard deviation and skewness) until the end

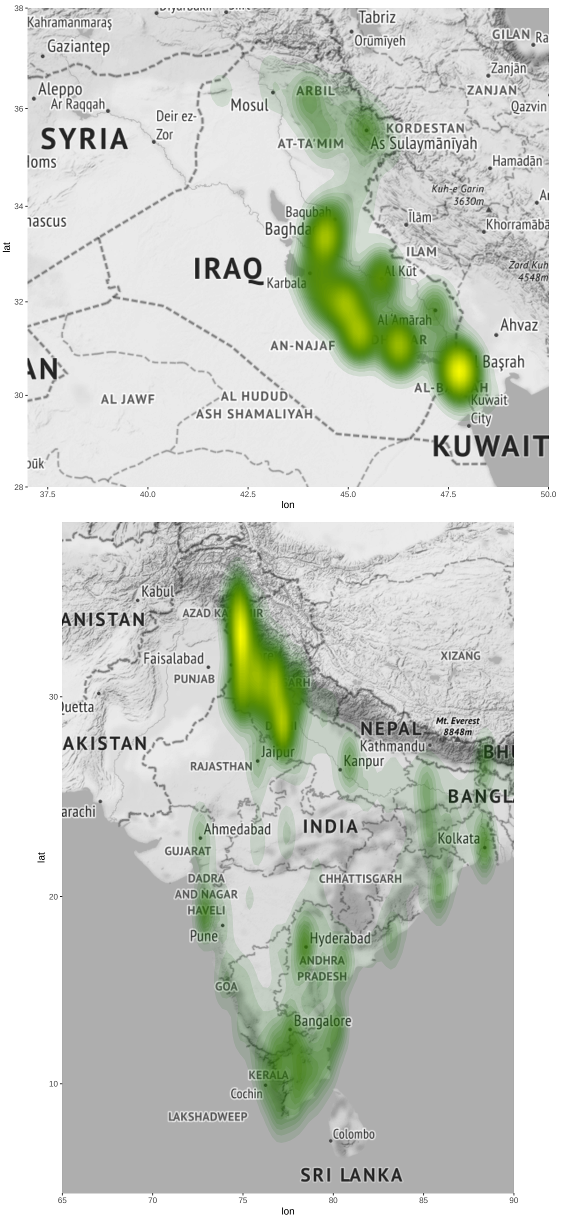


Figure 4. Localization of the protests in Iraq (top) and India (bottom) as density plots. Yellow areas represent a high density.

of the window and a trend for each indicator is derived (on the second half of each rolling window). The trend is given by Kendall τ rank correlation with values in the interval $[-1, 1]$. The Kendall τ is used in statistics to measure the ordinal association between two measured quantities: if $(x_1, y_1), \ldots, (x_n, y_n)$ is a set of observations, any pair of observation (x_i, y_i) and (x_j, y_j), where $i < j$, are said to be concordant if the sort order of (x_i, x_j) and (y_i, y_j) agrees, i.e., if either both $x_i > x_j$ and $y_i > y_j$ holds or both $x_i < x_j$ and $y_i < y_j$; otherwise they are said to be discordant. The Kendall τ is then defined as

$$\begin{aligned}\tau &= \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\binom{n}{2}} \\ &= \frac{2}{n(n-1)} \sum_{i<j} \operatorname{sgn}(x_i - x_j) \operatorname{sgn}(y_i - y_j).\end{aligned}$$

The procedure is sketched in Figure 5.

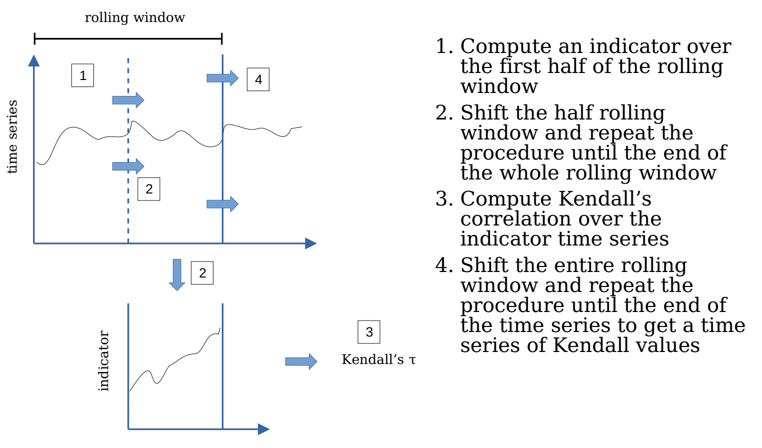


Figure 5. Method for generating a time series of Kendall values

Autocorrelation at-lag-1, standard deviation and skewness are respectively defined for a time series $(x_k)_{k=1}^n$ as

$$AC(1) = \frac{1}{n} \sum_{k=1}^{n-1} (x_k - \bar{x})(x_{k+1} - \bar{x}),$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2},$$

$$S = \frac{n}{(n-1)(n-2)} \sum_{k=1}^{n} \left(\frac{x_k - \bar{x}}{\sigma} \right)^3,$$

where $\bar{x}$ denotes the mean of (x_k). It is known from the theory of EWS that tipping points are preceded by an increasing trend of the Kendall τ-values [10]. This is the case for autocorrelation and, depending on the context, for standard deviation (in a critical slowing down scenario) or skewness (flickering) (see Figure 6).

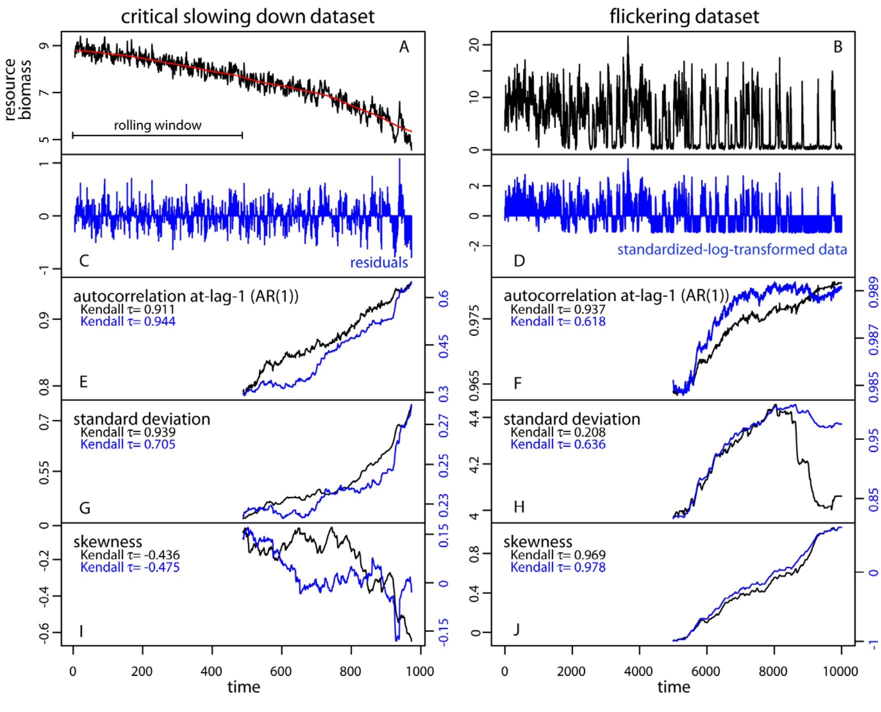


Figure 6. Kendall values for a set of indicators taken from [10]. The data represented in the top graphs A and B were synthetically generated based on an ecological model and it is known that a tipping point exists at the end of both time series. The blue time series C and D are obtained from the original ones using detrending methods (logarithmic transformation, Gaussian filtering). The graphs E to J represent indicators and their corresponding Kendall values. Here, a rolling window of length half of the total length of the time series is used.

Note that the list of used indicators is not restricted to the mentioned three ones. One can also use spectral indicators like the low frequency power spectrum (LFPS), the spectral or density ratio, perform a detrended fluctuation analysis (DFA), use the coefficient of variation, the kurtosis, the $W2$ index of Drake and Griffen, the conditional heteroskedasticity or BDS tests [10, 17, 11].

Although being interested in the mentioned events in Iraq and India we compute the trend of each indicator over the whole time series. We then build a significance test over the τ-values and finally conclude with a sensitivity analysis, investigating the effect of varying the length of the rolling windows.

The significance of the trends is tested following an idea used in [10]. A null hypothesis is simply formulated by stating that the trend estimates of the indicators are due to chance alone. We fit the transformed original time series by choosing the best ARMA(p, q) model based on the Akaike information criterion (AIC) for values of p and q not exceeding 3. We then use the model to generate a large number of surrogate datasets (say $n = 1\,000$) of the same length as the original dataset and next estimate, for each generated dataset, the trend of each indicator. It is so possible to compare the original trend with the distribution of the generated trends: we can determine a p-value as the proportion of τ-values in the generated distribution of trends that are larger than the trend of the original dataset.

7 RESULTS

7.1 Iraqi Protests

The trends, given as the time evolution of the Kendall values, are shown in Figure 7 b). We see a simultaneous increasing trend ending with relative high values before October 2019. This is especially the case for autocorrelation and standard deviation. However, each indicator can exhibit increasing trends independently of the other ones.

The time evolution of the trends expressed as p-values is shown in Figure 7 c) over the whole dataset. The analysis was performed using a rolling window size of 370 days, i.e., 20 % of the data.

In order to investigate the sensitivity of the results to the rolling window size we repeated the analysis before 1 October 2019 using different window sizes (see Figure 8). Window sizes between 350 and 400 days seem to be adequate choices but the results are relatively highly sensitive to this parameter.

The results shown also need to be put in the context of the political situation in Iraq. The one-year anniversary of the beginning of the protests was the occasion for protesters to take to the streets [2]. This is also visible in Figure 7.

7.2 Indian Protests

The same methods were applied to the Indian dataset. The results are shown in Figure 9. Here we used a rolling window of length 278 days corresponding to 15 % of the data.

It is noticeable that autocorrelation and standard deviation are significant in the Iraqi case, and standard deviation and skewness in the Indian case. Autocorrelation and skewness are furthermore significant at the beginning of December 2019 in the Iraqi case. The autocorrelation is also low in the Indian case but not significant. A similar phenomenon is also discernible for the skewness in the Iraqi case.

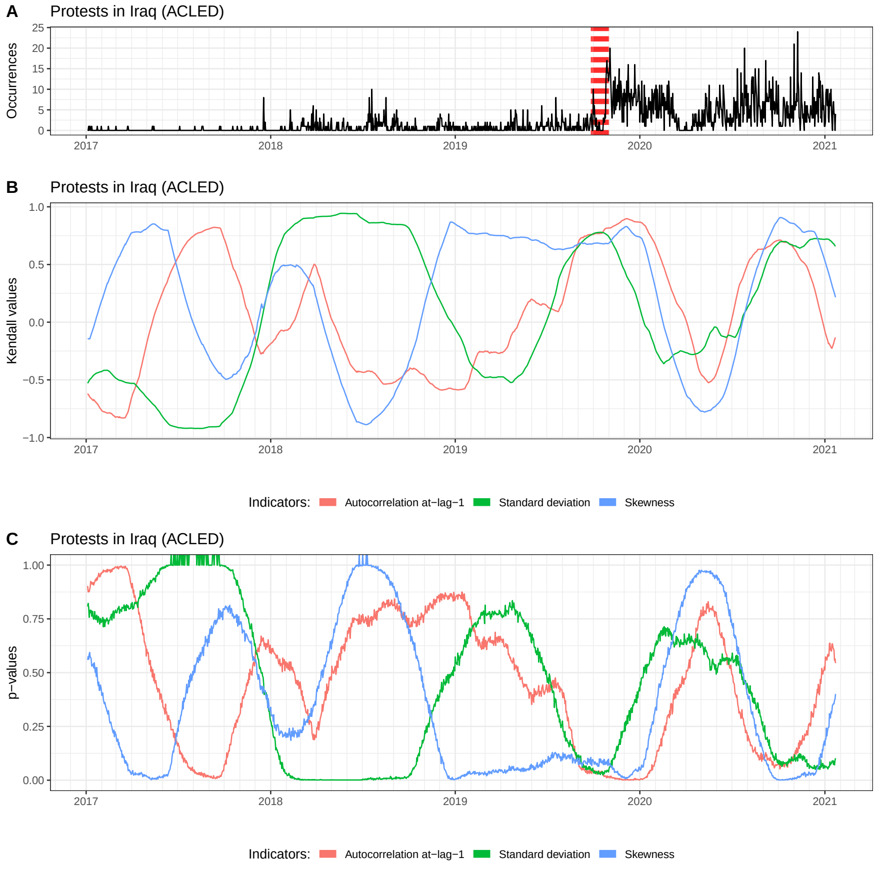
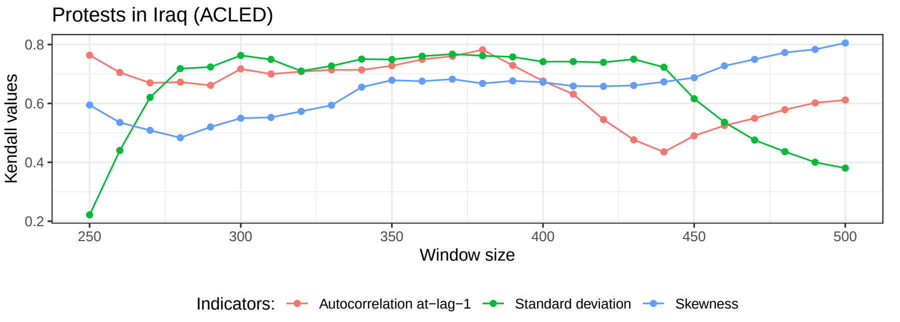


Figure 7. Daily protests in Iraq with indicator analysis. B) shows the time evolution of the Kendall τ-values of the indicators and C) the corresponding p-values. A) shows the time series with red vertical lines indicating time points where both autocorrelation and standard deviation are significant with p-values < 0.05. The first such detected time point is 27 September 2019. We also note that autocorrelation and skewness are significant with p-values < 0.01 at the beginning of December 2019. We used here $n = 1\,000$ surrogate datasets for the p-values.

We also give a sensitivity analysis based on the rolling window size in Figure 10.

8 DISCUSSION

We showed that EWS are useful methods to study the dynamics of OSINT time series with the use cases of the 2019 Iraqi and 2020 Indian farmers' protests. In the



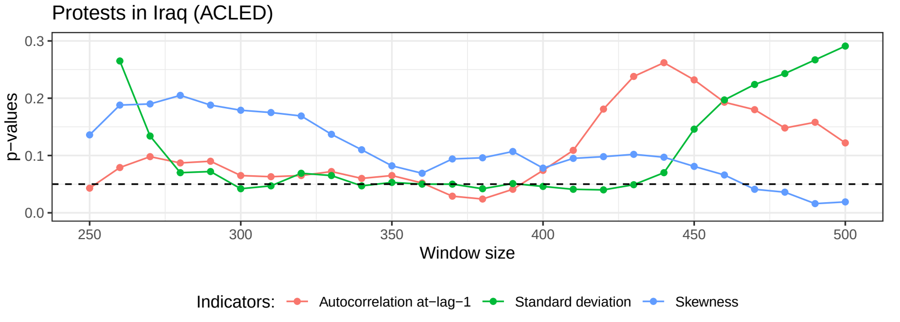


Figure 8. Sensitivity analysis. Windows of sizes between 250 and 500 are tested before 1 October 2019. We used here $n = 1\,000$ surrogate datasets for the p-values.

first case autocorrelation and standard deviation are significant while the second case is significant for standard deviation and skewness. Further analyses need to be performed on other datasets in order to validate the results presented here (e.g. Bulgaria [4], Chile [1], Belarus [3]). It would also be interesting to investigate other indicators, in particular the model-based EWS.

It would also be valuable to investigate which events can be anticipated with which indicators. The results in this work show that a combination of two of the three used indicators are good measures for these use cases but tipping points could lead to other types of dynamic changes as, e.g., a drop in the intensity.

In any case, the methods presented have a huge potential for building alarm systems on OSINT time series. The ACLED database does not only contain information about protests worldwide but also about riots, violence against civilians, explosions, battles, and strategic developments. This opens further aspects of predictive and causal analysis, regarding the escalating forms of social unrest. All ACLED events are moreover geolocalised. This also opens another field of research: the study of spatial EWS [12].

Protests are also increasingly preceded by activities on social networks. For example, the hashtag `#FarmersProtest` is highly widespread in Twitter and it could

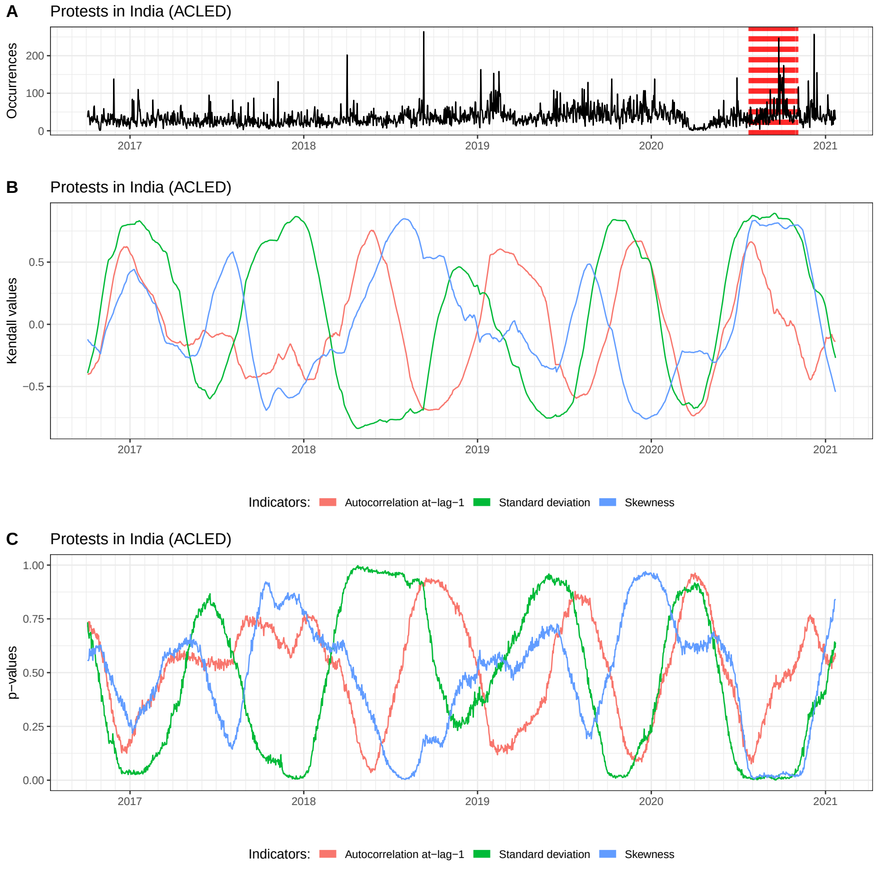


Figure 9. Daily protests in India with indicator analysis. B) shows the time evolution of the Kendall τ-values of the indicators and C) the corresponding p-values. A) shows the time series with red vertical lines indicating time points where both standard deviation and skewness are significant with p-values < 0.05. The first such detected time point is 24 July 2020. Note that both p-values are < 0.01 at the beginning of August 2020. We used here $n = 1\,000$ surrogate datasets for the p-values.

be interesting to track EWS on social media and link them to events in the real world.

9 CONCLUSION

In this work we could show that EWS do occur in OSINT time series using standard indicators and statistical testing. The dominant role of the standard deviation in this OSINT context was also exhibited.

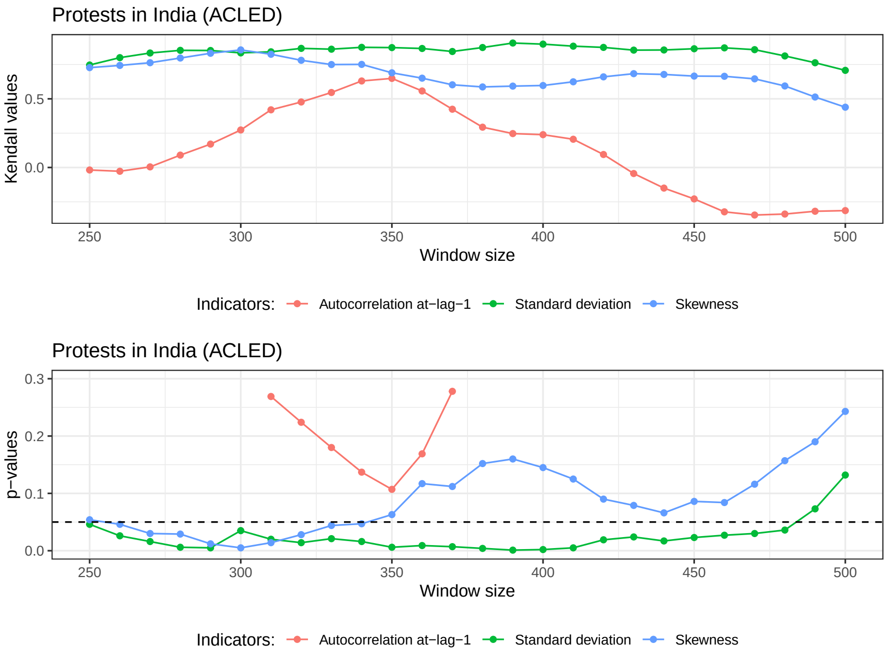


Figure 10. Sensitivity analysis. Windows of sizes between 250 and 500 are tested before 30 September 2020. We used here $n = 1\,000$ surrogate datasets for the p-values.

Methods for identifying EWS were first developed in ecological datasets using synthetic data, i.e., data generated from ecological dynamical models for which tipping points are known to exist. These methods are proven to be effective on real data in numerous fields of research where the dynamics is governed by physical laws.

In social sciences, the detection of unknown events is more difficult and cannot be always expected due to the influence of exogenous forces. One can nevertheless hope that protests and similar social systems satisfy common dynamical critical patterns as is shown in this work.

10 FUTURE WORK

Future work should focus on validating these results on more datasets. It is a priori unknown in which proportion such critical phenomena appear in the dynamics of protests and, in case of high predominance, what are the characteristics of such events? It would also be valuable to examine more indicators and select the most appropriate ones. These studies should also consider the sensitivity to the length of the rolling windows and robust methods are desirable.

Another interesting question arises from the origin of such protests. It is known that social media play nowadays a crucial role in the outbreak of protests and other real world events. It is therefore natural to seek similar phenomena in social networks and study the causality between social network events and the real ones.

REFERENCES

[1] 2019–2021 Chilean Protests. https://en.wikipedia.org/wiki/2019-2021_Chilean_protests.

[2] 2019–2021 Iraqi Protests. https://en.wikipedia.org/wiki/2019-2021_Iraqi_protests.

[3] 2020–2021 Belarusian Protests. https://en.wikipedia.org/wiki/2020-2021_Belarusian_protests.

[4] 2020–2021 Bulgarian Protests. https://en.wikipedia.org/wiki/2020-2021_Bulgarian_protests.

[5] 2020–2021 Indian Farmers' Protest. https://en.wikipedia.org/wiki/2020-2021_Indian_farmers%27_protest.

[6] The Armed Conflict Location and Event Data Project. https://acleddata.com.

[7] Early Warning Research Hub. https://acleddata.com/early-warning-research-hub/.

[8] Adams, R. P.—MacKay, D. J. C.: Bayesian Online Changepoint Detection. 2007, arXiv: 0710.3742.

[9] van den Burg, G. J. J.—Williams, C. K. I.: An Evaluation of Change Point Detection Algorithms. 2020, arXiv: 2003.06222v2.

[10] Dakos, V.—Carpenter, S. R.—Brock, W. A.—Ellison, A. M.—Guttal, V.—Ives, A. R.—Kéfi, S.—Livina, V.—Seekell, D. A.—van Nes, E. H.—Scheffer, M.: Methods for Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data. PLoS ONE, Vol. 7, 2021, No. 7, Art. No. e41010, doi: 10.1371/journal.pone.0041010.

[11] Drake, J. M.—Griffen, B. D.: Early Warning Signals of Extinction in Deteriorating Environments. Nature, Vol. 467, 2010, No. 7314, pp. 456–459, doi: 10.1038/nature09389.

[12] Kéfi, S.—Guttal, V.—Brock, W. A.—Carpenter, S. R.—Ellison, A. M.—Livina, V. N.—Seekell, D. A.—Scheffer, M.—van Nes, E. H.—Dakos, V.: Early Warning Signals of Ecological Transitions: Methods for Spatial Patterns. PLoS ONE, Vol. 9, 2014, No. 3, Art. No. e92097, doi: 10.1371/journal.pone.0092097.

[13] Kuehn, C.—Martens, E. A.—Romero, D. M.: Critical Transitions in Social Network Activity. 2018, arXiv: 1307.8250v2.

[14] Scheffer, M.: Critical Transitions in Nature and Society. Princeton Studies in Complexity, Princeton University Press, 2009, https://press.princeton.edu/books/paperback/9780691122045/critical-transitions-in-nature-and-society.

[15] Scheffer, M.—Bascompte, J.—Brock, W. A.—Brovkin, V.—Carpenter, S. R.—Dakos, V.—Held, H.—van Nes, E. H.—Rietkerk, M.—

Sugihara, G.: Early-Warning Signals for Critical Transitions. Nature, Vol. 461, 2009, No. 7260, pp. 53–59, doi: 10.1038/nature08227.

[16] Scheffer, M.—Carpenter, S.—Foley, J.—Folke, C.—Walker, B.: Catastrophic Shifts in Ecosystems. Nature, Vol. 413, 2001, No. 6856, pp. 591–596, doi: 10.1038/35098000.

[17] Slater, D.: Early Warning Signals of Tipping-Points in Blog Posts. The MITRE Corporation, 2012.

[18] Spielmann, K. A.—Peeples, M. A.—Glowacki, D. M.—Dugmore, A.: Early Warning Signals of Social Transformation: A Case Study from the US Southwest. PLoS ONE, Vol. 11, 2016, No. 10, Art. No. e0163685, doi: 10.1371/journal.pone.0163685.

[19] Wen, H.—Ciamarra, M. P.—Cheong, S. A.: How One Might Miss Early Warning Signals of Critical Transitions in Time Series Data: A Systematic Study of Two Major Currency Pairs. PLoS ONE, Vol. 13, 2018 No. 3, Art. No. e0191439, doi: 10.1371/journal.pone.0191439.

Étienne Voutaz is Data Scientist at the Cyber-Defence Campus of armasuisse. His fields of activity range from open source intelligence to augmented reality to create value from raw data from both an academic and practical perspective. He also has a long experience in cybersecurity management and holds a Ph.D. in mathematics (dynamical systems), an EMBA in ICT management and various ISACA certifications (CISA, CRISC, CGEIT).

Albert Blarer is Scientific Project Manager and Data Scientist in the Cyber Defence Campus at armasuisse Science + Technology. His main research subjects are predictive models and causal aspects in a variety of topics from the intelligence domain.