

# Detecting Global Events with Bayesian Changepoint Detection on Flight Data

**Alec Jacobs and Ian Kloo**

Department of Systems Engineering,  
United States Military Academy,  
West Point, NY 10996

Corresponding author's Email: [16alecj@gmail.com](mailto:16alecj@gmail.com)

**Author Note:** This paper was previously published and presented in the Donald R. Keith Memorial Capstone Conference at USMA in May of 2023. Alec Jacobs is a Second Lieutenant in the United States Army at Fort Novosel where he is currently at flight school. He graduated from the United States Military Academy in 2023 as a Systems Engineering Major with a focus in Cyber Security. Mr. Ian Kloo is an Assistant Professor and Data Scientist within the Department of Systems Engineering and serves as the advisor. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

**Abstract:** As a means of enhancing the intelligence community's ability to be prepared and prompt in the detection of global events, we present a Bayesian changepoint detection methodology to detect anomalous activity in open-source flight data. We demonstrate a computationally inexpensive methodology to monitor a geographic location for changes in flight activity over time and demonstrate its utility in a case study of the 2022 Russian invasion of Ukraine. Beyond flight data, the methodology described in this paper shows the potential to generate near real-time situational awareness using open-source time series data streams. This paper was previously published and presented in the Donald R. Keith Memorial Capstone Conference at USMA in May of 2023.

*Keywords:* Open-Source Intelligence, Flight Data, Time Series Analysis, Bayesian Changepoint Detection

## 1. Introduction

Recent events like the Chinese weather balloons over the United States highlight the importance of intelligence gathering as a geopolitical tool and show the risks that countries will take to collect data (Nakashima, Harris, & Samenow, n.d.). Fortunately, the evolution of the global information environment provides ample opportunities for intelligence analysis that does not violate international law. This paper demonstrates a methodology to detect global events using publicly available flight data. The presented methodology provides a computationally inexpensive method to monitor the entire world for major events that are correlated with changes in flight activity.

The authors developed the concept for this project after seeing a social media account run by a 20-year-old college student, Jack Sweeney. Sweeney's early work tracked Elon Musk's private jet and published his location as he moved around the world (Kay & Rains, n.d.). From there, the project evolved to track other celebrities and eventually the private jets owned by Russian oligarchs. Sweeney posted real-time oligarch locations on the Twitter account, @RUOligarchJets, which gained a significant following.

Instead of focusing on the individual planes, the authors decided to explore the idea that aggregated flight data could be used as a way to identify major geopolitical events. Sweeney's project showed that many oligarchs moved their jets out of Russia at the beginning of the invasion of Ukraine and it is possible that other major global events may correlate with changes in flight activity.

In the remainder of this paper, the authors will first describe the purpose of this project and provide some background into open-source intelligence (OSINT) in general. Then, the team will provide a literature review of the technologies leveraged in this study: time series analysis and Bayesian changepoint detection. Next, we describe the methodology and show how it can

be applied in a case study on the Russian invasion of Ukraine. Finally, the authors discuss some of the limitations of this work and propose the next steps to further the research to leverage flight data as an OSINT data stream.

## 2. Problem Statement

The purpose of this project is to create a proof-of-concept open-source intelligence tool using publicly available flight data. We hypothesize that we will be able to detect significant international events by finding changepoints in geographically specific flight traffic.

## 3. Current Process

Intelligence is partitioned into six primary divisions: human intelligence, geospatial intelligence, measure and signature intelligence, financial intelligence, signals intelligence, and open-source intelligence (Unver, 2018). This project focuses only on OSINT, though it could certainly be enhanced by data streams from the other intelligence disciplines. Unlike many of the other types of intelligence, OSINT leverages publicly available (unclassified) information to generate actionable intelligence.

In recent years, several OSINT projects have focused on social media platforms as data sources (Avellaneda-Ruiz, Bagley, Brown, Manzonelli, & Kloof, 2021; Yeboah-Ofori & Brimicombe, 2018; Shcherbakova & Gubina, 2022), but there are many data sources beyond social media that remain unused. Inexpensive computing resources and an increased appetite for data science have given rise to greater access to public application programming interfaces (APIs). This project will focus on the OpenSky Network's API which provides current and historical data on global flights (OpenSky-Network, 2023). The project team is not aware of any existing tools that utilize public flight data as a form of intelligence.

## 4. Literature Review

As mentioned in the previous section, the authors are currently unaware of similar OSINT applications using flight data. This specific application will leverage existing time series methods as well as Bayesian changepoint detection to create actionable intelligence and situational awareness. Before describing the methodology, we present a brief literature review on these topics to orient our project within the existing research.

### 4.1. Time Series Methods

A time series is simply a list of numbers representing some data along with the sequential times those data points were recorded (Hyndman & Athanasopoulos, 2018). For this application, the team's focus is exclusively on univariate, non-stationary time series. Specifically, the team found that flight data contains trend (in all tested cases) and seasonality (in some tested cases). Hyndman describes several methods for creating stationary time series, the simplest being first-order differencing. This method begins at period  $n + 1$  and subtracts period  $n$ , dropping the first data point (Hyndman & Athanasopoulos, 2018). The resulting time series is now a measure of the 1-period change in the data instead of the raw data itself. Absent large-scale seasonality, the differenced data creates a detrended stationary time series that is centered around mean zero with random variation.

Hyndman's work describes many time series methods that take stationary data as inputs, but his work primarily focuses on forecasting. This project is instead looking for anomalous activity in the historical data, so the authors will pass the time series data to a Bayesian changepoint detection methodology.

### 4.2. Bayesian Methods

Bayesian changepoint methods are underpinned by Bayes' theorem, shown in Equation 1:

$$\rho(M | D) = \frac{\rho(D | M)\rho(M)}{\rho(D)} \quad (1)$$

The probability of an initial belief model,  $M$ , given the new data,  $D$ , is equal to the probability of seeing this data given the initial belief, times the probability of the initial belief divided by the probability of seeing this data (Ivezić, Connolly, VanderPlas, & Gray, 2014). In plain English, Bayes' Theorem takes some prior belief and updates it based on some observed data.

A changepoint technique described by Barry and Hartigan (1993) leverages Bayesian statistics by modeling a series of data as if it is generated by a given distribution with a set of parameters that are learned from the data. The model then estimates the likelihood that each data point was generated given the supposed distribution and parameters. Any data points that appear to be generated by another set of parameters signal a likely changepoint. The authors use the Bayesian Analysis of Changepoints (BCP) R package which is a computationally efficient implementation of Barry and Hartigan's work (Erdman & Emerson, 2007). The package utilizes both a multivariate and linear regression changepoint analysis to partition the data into blocks, with a mean assumed to be constant between blocks. This assumption allows for the data to be treated as independent and identically distributed, which allows deviations to be considered anomalies. (Erdman & Emerson, 2007)

Sharma, Swayne, and Obimbo (2016) conducted an alternative comparison for multiple changepoint detection methods to include BCP, Wild Binary Segmentation (WBS), E-Agglomerative algorithm for changepoint, and Iterative Robust Detection. The authors of this study suggest that WBS is the best model for most use cases, however this package does not provide the human-readable probabilistic output that BCP generates. The main drawback to BCP that Sharma et. al. describe is a greater potential for false positives when compared to the other methods (Sharma, Swayne, & Obimbo, 2016). For this project, the team is willing to accept a high-recall approach at the cost of precision especially given BCP's highly interpretable output.

## 5. Methodology



Figure 1: A High-level methodology for an OSINT tool using flight frequency data.

Figure 1 illustrates a high-level view of the methodology. The first step is to collect (or continuously monitoring) flight data from OpenSky API. Next, we process the data and run a Bayesian changepoint analysis to detect potential anomalies. Any anomalies create alerts for intelligence analysts to further investigate, eventually providing actionable intelligence. The remainder of this section will describe the data collection, data transformation, and changepoint analysis in greater detail. The authors will not attempt to describe the steps after a potential anomaly is detected as these will be highly dependent on doctrine and organizational differences of the users.

### 5.1. Flight Data Collection and Processing

The OpenSky Network's API provides every flight departing from any airport in a specified time period (OpenSkyNetwork, 2023). The project used the openSkies R package for this project to avoid writing HTTP requests (Ayala, Ayala, Ruiz, Sellés, & Vidal, 2021). The openSkies package allows for robust data access without the need to authenticate to the API; however, signing up for a free account increases the speed of data collection by increasing short-term rate limits.

The openSkies output is in a list format where each value is a set of text vectors describing an individual flight. The authors parsed these text lists, resulting in vector of dates (with time) such that there is an entry for every flight. They then used base R's time series (ts) function to aggregate the data to various intervals including hour, day, and week (R Core Team, 2022).

The methodology uses OpenSky as a data source, but there is nothing about the work that limits it to using this data source alone. OpenSky collects Automatic Dependent Surveillance–Broadcast (ADS-B) data which is a protocol that most

planes are required to transmit globally and is accessible to anyone with an appropriate receiver (OpenSky-Network, 2023). It is important that any potential users of this methodology consult OpenSky's terms of use to determine if they are permitted to use the API for their specific purposes. While OpenSky is a convenient service, we anticipate being able to access ADS-B data by other means if necessary.

## 5.2. Data Transformation and Changepoint Detection

After parsing the data into a time series format, it is important to detrend the data to create a stationary time series (as discussed in the Literature Review section). The study team found first-order differencing is sufficient to detrend all of the flight data we tested. Figure 2 shows a set of flight data that exhibits several different trends and the subsequent detrended series after taking the first-order difference.

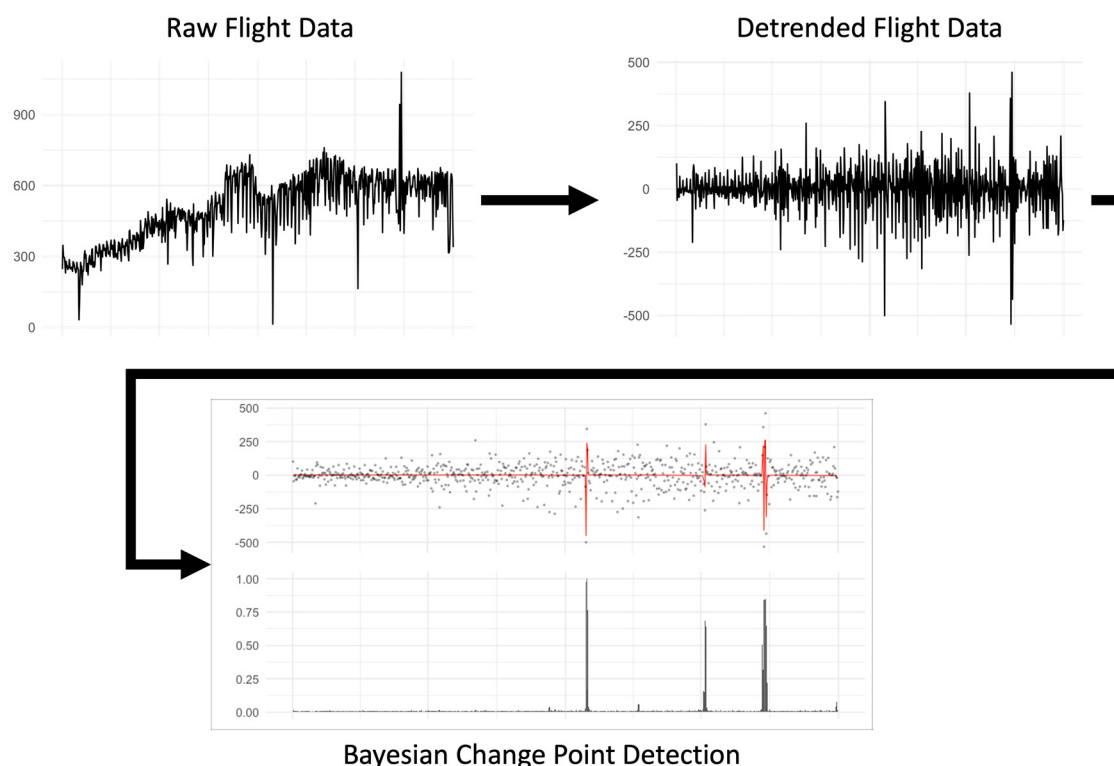


Figure 2: Demonstration of the data transformation process from raw data to detrended data, and finally changepoint detection. The unit of analysis for the top plots is the number of flights departing the four Moscow Airports per time period while the y axis of the bottom plot represents the probability of a changepoint.

The detrended data is then passed to the Bayesian changepoint algorithm using the BCP R package (as described in the Literature Review section) (Erdman & Emerson, 2007). BCP takes two parameters:  $p_0$  is the parameter on the prior on changepoint probabilities and  $w_0$  is the parameter on the prior of the signal-to-noise ratio (Erdman & Emerson, 2007). The authors found that the default values (0.2 for both parameters) were sufficient for all of the test cases and the package authors suggest that modifying them is not often necessary. The study team tested several additional parameter values and found that only extreme values produced different results.

BCP output is shown in the bottom row of Figure 2. The red line shows the posterior mean of the data while the black dots show the actual data points. The black bars on the bottom of the visualization show the posterior probability that a change occurred at each location. There are three likely changepoints in the plot shown in Figure 2.

## 6. Case Study: Russian Invasion of Ukraine

To demonstrate the methodology, the authors present a case study on the Russian invasion of Ukraine that began in 2022. The authors selected four major airports around Moscow: Sheremetyevo, Moscow Domodedovo, Zhukovsky, and Vnukovo and gathered daily flight traffic from January 1, 2021 to January 1, 2023. After detrending the data by taking the first-order difference, the team ran Bayesian changepoint detection. The results are shown in Figure 3.

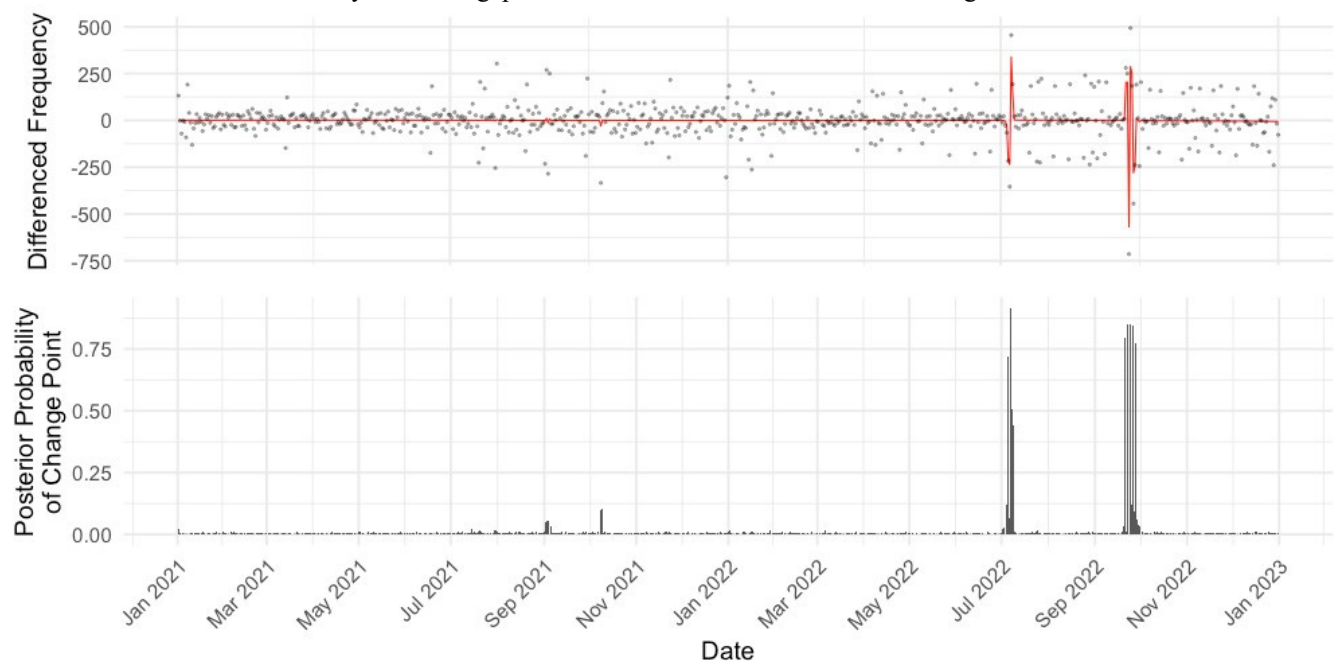


Figure 3: Model output showing the posterior probabilities of changepoints in Russian flight traffic during the invasion of Ukraine. Two main changepoints are evident at significant dates pertaining to mobilization announcements.

The Russian invasion officially began on February 24, 2022 (Reuters, 2022a), and the authors anticipated an increase in flight traffic around that date; however, there are no changepoints in early 2022. A similar analysis on the hourly and weekly time scales found hourly to be overly granular (resulting in many false positives) and weekly to be smoothed to the point that changepoints cannot be detected. It appears that the Russian invasion did not result in a noticeable change in air traffic from our selected airports.

There are clear changepoints in early July and late September 2022. Recall that the top pane of Figure 3 shows first-order differenced data, so the interpretation of these findings is that the daily difference in flight traffic changed in both early July and late September. Both of these changepoints correspond to key announcements pertaining to troop mobilization efforts that we suspect are causally related the change in flight traffic.

In July 2022, Russia began two mobilization efforts. First, the paramilitary organization, The Wagner Group, started recruiting prisoners (Hird, Stepanenko, Mappers, & Kagan, 2022a; Reuters, 2023). Second, there was a movement some journalists referred to as the “covert mobilization” that saw military and paramilitary organizations dramatically increasing their recruitment efforts and offering new incentives to new recruits (Hird et al., 2022a; Reuters, 2022b). These two efforts were

the first major mobilization efforts in a war that Vladimir Putin still maintained would not require additional manpower (Reuters, 2022b). The authors hypothesize that, in combination, these mobilization efforts created concern among Russian citizens who sought to leave the country. Additionally, it is possible that the Russian government engaged in efforts to remove opportunities for people to leave the country in anticipation of the concern around potential future mobilizations.

The mobilization effort in September 2022 was much more overt. In a September 21st speech, Vladimir Putin announced a formal partial mobilization of Russia in support of the war effort (Hird, Stepanenko, Mappers, & Kagan, 2022b). This was the first admission from Putin that the war would require mobilization beyond the existing military force, so it was understandably met with surprise by some Russian citizens. There were reports after the speech that huge amounts of Russians left (or tried to leave) the country (Maynes, 2022), but Russian official channels disputed this claim (Zakir-Hussain, 2022). The project's data supports the idea that there were major changes in flight patterns in the days immediately following the speech. As with the July change points, the authors observe both increases and decreases in flight traffic. Taken in conjunction with Russia's insistence that people were not trying to leave the country, the team hypothesizes that the government may have taken action to stem the flow of citizens out of Russia in response to an increase in flight travel.

## 7. Limitations and Future Work

The Russian case study shows that this methodology can identify some major geopolitical events, but those events must be correlated with changes in air travel. The authors anticipated that the start of the Russian invasion would impact air travel and were surprised to find that it did not. The team believed Russian citizens would be wary of conscription or adverse effects such as impacts to business and the potential to travel in the future, affecting all citizens, not just the wealthy. The study team is also interested in looking into other countries involved, like Ukraine, and other surrounding countries, like Poland, as well to see their impact. The team lacks the intuition to hypothesize the types of events that will likely be detected by the methodology and recommend further study into known major geopolitical events to determine the types of things this methodology can be expected to find.

A key limitation in this study and any future application of the methodology is the completeness of the flight data that is used as an input. OpenSky presents a highly accessible database of flight information, but it is possible that regulatory changes will impact the number of aircraft using the protocols it detects. The study team recommends investigating additional open-source flight data, potentially ensembling the data collection with OpenSky to create a more robust input data set.

Additionally, the model currently relies on univariate time series analysis. It would be valuable to explore additional features that may impact air travel over time. It is possible that including these variables could explain (and control for) the trends we saw in the data, potentially eliminating the need to difference the data. Removing the detrending step would make the final results much more interpretable. The BCP package supports multivariate regression inputs, so this seems like a natural next step for the project.

Finally, while the authors believe BCP is a good fit for this project, they recommend exploring additional change detection algorithms. Daily flight data is small in scale even over multiple years and it would be computationally reasonable to run several state-of-the-art change point models. The study team recommends exploring whether ensembling these models results in more accurate change point predictions.

## 8. Conclusion

This paper demonstrates that open-source flight data can be used to detect major geopolitical events as they occur using Bayesian change point detection. Although it can only detect events that are correlated with changes in air travel, it demonstrated that flight data should be considered as a part of a comprehensive set of OSINT data streams. In addition to highlighting the potential value of flight data, this paper shows a useful methodology for processing noisy time series data with trend to ultimately detect likely change points with human-readable output. This methodology could be applied to many similar time series data streams with minor modifications. Finally, this paper provides a road map for future study into the ways flight data could aid the OSINT process.

## References

- Avellaneda-Ruiz, A., Bagley, W., Brown, T., Manzonelli, N., & Kloo, I. (2021). Leveraging publicly available information to analyze information operations.
- Ayala, R., Ayala, D., Ruiz, D., Sellés, A., & Vidal, L. S. (2021). openskies: Retrieval, analysis and visualization of air traffic data [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=openSkies> (R package version 1.1.6)
- Barry, D., & Hartigan, J. A. (1993). A bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421), 309–319.
- Beaulieu, C., Chen, J., & Sarmiento, J. L. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 370(1962), 1228–1249. Retrieved 2023-03-28, from <http://www.jstor.org/stable/41348440>
- Cooper, P. (2017). Aviation cybersecurity: Finding lift, minimizing drag (Tech. Rep.). Atlantic Council. Retrieved 2023-03-28, from <http://www.jstor.org/stable/resrep16767.12>
- Erdman, C., & Emerson, J. W. (2007). bcp: An R package for performing a bayesian analysis of change point problems. *Journal of Statistical Software*, 23(3), 1–13. Retrieved from <http://www.jstatsoft.org/v23/i03/>
- Hird, K., Stepanenko, K., Mappers, G., & Kagan, F. W. (2022a, Jul). Russian offensive campaign assessment, july 10. Institute for the Study of War. Retrieved from <https://www.understandingwar.org/backgrounder/russian-offensive-campaign-assessment-july-10>
- Hird, K., Stepanenko, K., Mappers, G., & Kagan, F. W. (2022b, Sep). Russian offensive campaign assessment, september 21. Institute for the Study of War. Retrieved from <https://www.understandingwar.org/backgrounder/russian-offensive-campaign-assessment-september-21>
- Hyndman, R. J., & Athanasopoulos, G. (2018). In *Forecasting: Principles and practice* (2nd ed.). OTexts: Melbourne, Australia.
- Ine, B. E. (2011). Hidden in plain sight: The ever-increasing use of open source intelligence. *American Intelligence Journal*, 29(2), 141–144. Retrieved 2023-03-24, from <http://www.jstor.org/stable/26201962>
- Ivezić, Z., Connolly, A. J., VanderPlas, J. T., & Gray, A. (2014). Bayesian statistical inference. In *Statistics, data mining, and machine learning in astronomy: A practical python guide for the analysis of survey data* (STU - Student edition ed., pp. 175–246). Princeton University Press. Retrieved 2022-11-17, from <http://www.jstor.org/stable/j.ctt4cgbdj.8>
- Kay, G., & Rains, T. (n.d.). The college student who tracks elon musk’s private jet is launching his own flight-tracking website. *Business Insider*. Retrieved from <https://www.businessinsider.com/student-who-tracks-elon-musk-jet-launching-own-website-2023-2>
- Maynes, C. (2022, Sep). Russians are protesting and fleeing the country as putin orders a draft for ukraine. *NPR*. Retrieved from <https://www.npr.org/2022/09/23/1124678888/russia-ukraine-military-draft-protests-flight>
- Nakashima, E., Harris, S., & Samenow, J. (n.d.). U.s. tracked china spy balloon from launch on hainan island along unusual path. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/national-security/2023/02/14/china-spy-balloon-path-tracking-weather/>
- OpenSky-Network. (2023). Network opensky explorer emergency alerts coverage amp; facts receiver ranking vhf/voice feeding. *The OpenSky Network*. Retrieved from <https://opensky-network.org/>
- R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Reuters. (2022a, Feb). Russia invades ukraine: Political, military leaders’ reactions. Thomson Reuters. Retrieved from <https://www.reuters.com/markets/europe/leaders-key-quotes-after-russia-invades-ukraine-2022-02-24/>
- Reuters. (2022b, Jun). Russia is conducting covert mobilisation campaign, ukraine spy chief says. Thomson Reuters. Retrieved from <https://www.reuters.com/world/europe/russia-is-conducting-covert-mobilisation-campaign-ukraine-spy-chief-2022-06-25/>
- Reuters. (2023, Feb). Russia’s wagner mercenaries halt prisoner recruitment campaign - prigozhin. Thomson Reuters. Retrieved from <https://www.reuters.com/world/europe/russias-wagner-mercenaries-halt-prisoner-recruitment-campaign-founder-prigozhin-2023-02-09/>
- Sharma, S., Swayne, D., & Obimbo, C. (2016). Trend analysis and change point techniques: A survey. In *Energy, ecology and environment* (pp. 123–130).
- Shcherbakova, K., & Gubina, D. (2022). Application of open source intelligence (osint) technology for social media analysis. *Language in the field of professional communication.—Yekaterinburg, 2022*, 106–113.

- Unver, H. A. (2018). Digital open source intelligence and international security: A primer (Tech. Rep.). Centre for Economics and Foreign Policy Studies. Retrieved 2023-03-24, from <http://www.jstor.org/stable/resrep21048>
- Yeboah-Ofori, A., & Brimicombe, A. (2018). Cyber intelligence and osint: developing mitigation techniques against cybercrime threats on social media. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, 7(1), 87–98.
- Zakir-Hussain, M. (2022, Oct). Moscow denies 700,000 have fled since call-up – but doesn't have 'exact figure'. *Independent Digital News and Media*. Retrieved from <https://www.independent.co.uk/news/world/europe/ukraine-russia-mobilisation-peskov-b2196892.html>