# Unfolding the crisis in Sudan through Remote Sensing and Google Earth Engine.

SAMEERA SIDDIQUI

This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

Module title and code: Dissertation 23/24 CASA 0010

Supervisor: Dr. Ollie Ballinger

Course : MSc. Urban Spatial Science

Student number: 23140058

Word Count:

# Table of Content

[Title Page 1](#_Toc174387251)

[Uncovering the state of Sudan civil war through remote sensing and google earth engine 1](#_Toc174387252)

[Table of Content 2](#_Toc174387253)

[Acknowledgements 3](#_Toc174387254)

[Abstract 5](#_Toc174387255)

[List of figures 6](#_Toc174387256)

[List of tables 7](#_Toc174387257)

[Abbreviations 8](#_Toc174387258)

[1. Introduction 9](#_Toc174387259)

[2. Literature Review 13](#_Toc174387260)

[3. Study area and data 17](#_Toc174387261)

[4. Methods 21](#_Toc174387262)

[5. Results and discussion 32](#_Toc174387263)

# List of figures

Figure 1: Smoke rises over buildings after aerial bombardment, during clashes between the RSF and the army in Khartoum North, Sudan, May 1, 2023

Figure 2 : Population density Map of Sudan

Figure 3: MODIS Fire ground Observation

Figure 4: Grid Formation of Sudan in GEE.

Figure 5: ACLED Events and Sub-event types (ACLED, 2023)

Figure 6: Trend of the Input Data

Figure 7: Map of Sudan for Year 2003, 2004 and 2023

Figure 8: Scatter Plot (Before Log and After Log)

Figure 9: KDE Graph of Z-score

# List of tables

Table 1: Panel Data frame

Table 2: Geo Data frame

Table 3: Confusion Matrix

Table 4: Model Comparision

# Acknowledgements

Ha, a sigh of relief I finally manage to finish my dissertation and what I envisioned to do. I have put all my learnings into this research, and it has finally took a shape. All the hard work of past 4 months has come to an end.

I would like to say thanks to all my teachers from Bartlett CASA. Last year in September when I came here, I could not have imagined myself learning so much and putting this much effort into the course.

A special thanks goes to my supervisor Dr. Ollie Ballinger who guided me all through the way with his immense support and constructive feedback. He answered to all my confusion and queries very patiently. Another special mention is to my partner Asif without whom I could not have finished this research. Whenever I felt stuck or could not think beyond, he helped me to look through and helped me with ideas and motivated me to do better and my 5-year-old boy Azaad though he did not help in writing but he sat there patiently watched me writing.

Lastly, I would like to say thank you to all my friends who believed in me and my vision. I want to dedicate this study to all those people who are suffering in war all over the world. May peace prevails.

In the end I would take responsibility for this research if there were any flaws or errors.

# Abstract

With the advancement of technology, particularly in social media platforms and use of remote sensing tools, tracking wars and damage caused all over the world has become significantly easier for providing relief and aid to the survivors.

Remotely sensed earth observation data has been used for informing decisions on environmental hazards arising from changing climate in urban areas, green space coverage, pollution studies, flooding, disaster response and many other applications. Here it is used for monitoring conflicts. The main aim here is to create an open monitoring system which tracks fires and conflict event and can also can differentiate between anomalous fires which are supposed to be conflict induced.

According to the literature number of times fire is used as the main weapon in conflicts. Therefore many events link to violence often lead to outbreak of fires and it can take the form of intentional acts of arson such as the burning of buildings, accidental ignition of flammable materials due to the use of explosives, burning debris during reconstruction efforts, or the destruction of evidence related to potential war crimes, among other examples and then there are natural fires like forest fire, or man-made seasonal agricultural fires. Most of the researchers working on fire data face difficulties in differentiating between the conflict fires and regular fires.

Many times false fires are detected and there is lot of mismatch between the conflicts reported and fires spotted. Therefore, this study is focused on Sudan crisis which started in April 2023 by two power groups for political reasons and still it is going on. This is the conflict which no one is talking about having less media coverage and attention it’s important to acknowledge this conflict by using remote sensing satellite imagery and google earth engine cloud platform. This work focuses on developing an open monitoring system using FIRMS fire data and ACLED conflict data for finding any relation between conflict and fires. The study uses panel dataset which was created by dividing Sudan into grids. Each grid gives the count of fires and ACLED event for 24 years from 2000- 2023. It also calculate the z-score of post conflict and pre conflict and by using the threshold value of z-score a new column of conflict induced fires can be calculated which further can be used to perform panel regression which gives some relation between fires and ACLED data.

# Abbreviations

ACLED- Armed Conflict Location & Event Data Project

AFP- Agence France press

AP- Associated press

CIR- Centre for Information Resilience

EPSG-European Petroleum Survey Group

FIRMS – Fire Information for Resource Management System

MODIS- Moderate Resolution Imaging Spectroradiometer

NRT- Near Real Time

RSF – Rapid Support Forces

SAF- Sudan Armed Forces

VIIRS- Visible Infrared Imaging Radiometer Suite

# Introduction

In the annals of history, wars and conflicts have cast a somber shadow, leaving behind a trail of sorrow and tragedy. From the ongoing struggles in Gaza, Ukraine, Sudan, Myanmar, Syria, and Ethiopia, the current situation paints a grim picture of uncertainty and unrest.

The internet and smartphones enable the sharing of information through various online platforms and applications. These include Twitter, Facebook, YouTube, Telegram, and discussion boards like Reddit. A crucial aspect of this information sharing is the use of remote sensing data. Remote sensing technology helps track and analyze the links between conflicts and environmental damage, providing valuable insights into the environmental impact of human activities and natural events.

There’s a humanitarian crisis happening since past year in Sudan which has been overshadowed by Gaza and Ukraine and it’s getting worse. (Wim Zwijnenburg and Ballinger, 2023; Kristof, 2024).

1.1 Research Background

The Sudanese civil war centers around a power struggle between two major groups: the Sudan Armed Forces (SAF) led by general Abdel Fattah al-Burhan, and the paramilitary Rapid Support Forces (RSF) commanded by General Mohamed Hamdan Dagalo, also known by ‘Hemedti’. (Ali, T., 2024)

Since, the conflict broke out on 15th April 2023, it has resulted in the loss of more than 18,800 people and over 33,000 injured, according to humanitarian partners. Over 10 million people have fled their homes, and this includes more than 5 million children and over 2 million people who have crossed into neighboring countries. (OCHA, 2024)

The country is facing extreme shortages of food, water, medicine, and fuel and nearly 18 million people are facing acute food insecurity and 5 million of them at emergency levels. Before the current conflict, Sudan had been grappling with violence and displacement since the onset of the Darfur crisis in 2003. (UNHCR ,2024)

‌A large black smoke billowing from a building

Description automatically generated

Figure 1: Smoke rises over buildings after aerial bombardment, during clashes between the paramilitary Rapid Support Forces and the army in Khartoum North, Sudan, May 1, 2023. REUTERS/Mohamed Nureldin Abdallah. (Abdelaziz, K. and Lewis, A., 2023)

According to investigators from the Centre for Information Resilience’s (CIR) Sudan Witness project, fire constitute to devastate villages and settlements in western Sudan. (CIR, 2024) The reports of fighting or airstrikes coinciding with clusters of fires indicates that fire is used as a weapon of war.

* The Geo-political Aspect

The ongoing civil war in Sudan holds immense geopolitical significance due to the country's strategic location near the Suez Canal in the Horn of Africa, a vital corridor for global trade. Sudan's abundant oil and gold reserves have drawn the interest of various external actors aiming to advance their own strategic interests, further intensifying the conflict. (Vardhan, A., 2023)

The RSF profit from gold trade with Russia and the UAE. The UAE refines most of Sudan’s gold, sourced from RSF leader Hemedti’s military business. Russia, via the Wagner Group, trains the RSF in exchange for gold access. (Doxsee, 2023) Egypt remains neutral due to its economic ties with the UAE. (Rickett, 2024; ADF, 2024; Trad, 2023) This intricate web of alliances and interests positions the Sudanese civil war as a potential proxy battleground, with regional and global powers maneuvering to secure their stakes. The U.S. has clearly stated that all parties supplying weapons should stop, as these actions are exacerbating the conflict and endangering Sudanese lives. (Nations, U.S.M. to the U., 2024)

* + The Role of Media – GAP in Coverage

Media coverage of the Sudan conflict has been limited compared to other global crises like the Israel-Gaza conflict and the Ukraine war, which receive more attention due to their direct implications for global security. Restricted access to conflict zones in Sudan, including a ban on foreign media until recently, has also hindered reporting. Major outlets such as The Washington Post and The New York Times offer limited coverage, while agencies like AP, AFP, and Reuters continue to report on the violence and humanitarian issues.

Remote sensing methods come to rescue even if the access was there as the war is happening in entire country. It can monitor areas that are difficult to cover on the ground.

* 1. Research Question and Objective
* How to identify conflict induced fires in Sudan by using FIRMS data and ACLED conflict data?
* Compare panel data regression models and suggest best modelling approach for prediction of conflicts using FIRMS data.

The purpose of this research is to identify conflict induced fires using FIRMS data and ACLED conflict data and to explore the relationship between them. The approach can be used to develop real-time conflict monitoring system based on satellite imagery for regions which have less media coverage and facing similar kind of civil war involving fire as the main weapon.

# Literature Review

This section introduces the background and motivation for the work carried out for this report, which have been introduced above in the form of similar studies done using remote sensing in war impacted regions where some gaps are identified and how with the methods explored in the research can improve the ways of identifying war related fires.

Conflict research can be divided into four main categories followed by the temporal lag for which each effect requires to become visible. For example, conflict-induced structural damage from bomb detonations or fire is generally a very immediate impact, occurring within minutes to hours. Other effects such as environmental damage (hours to days), population movement (days to months), and land-cover/land-use change (months to years) take longer to materialize. Although there is some overlap between different conflict effects, this typology provides a useful way to organize the current body of research. The three-paper discussed here discusses all these categories.

Bromley, L.(2010) paper had covered the conflict of 2003 in Darfur, Sudan using Moderate Resolution Imaging Spectroradiometer (MODIS) fire data to detect fires and burned area. The paper was based on the finding that using remote sensing for violent conflicts produce heat in the form of fires, incendiary bombs, etc., and that satellite-based instruments can detect those heat events. During that time, it was very difficult to know the exact location of the conflict as to gather ground truth data was not possible and the restrictions from government which allowed very minimal contingents of African union to move so the information was sparse that where the attacks took place. It was challenging but one thing which worked was the involvement of social media platform which played an important role in covering this civil war where foreign journalist was banned but organization like CIR which were involved in investigating, actively monitoring and documenting and verify human rights interferences and conflict dynamics on the ground. (Centre for Informati., 2023)

The vast majority of studies on Bromley, L.(2010) paper have been qualitative. The data used only covered major towns where journalist had greater access to refugees leaving the other parts which might have been affected also and the dataset used does not differentiate between the burned settlements from otherwise destroyed settlement and neither indicates the date of destruction. The paper lacks clarity with MODIS fire detection system which was able to sample only larger fires.

Looking at all the limitations the following research aims to develop a methodology which uses Fire Information for Resource Management System (FIRMS) fire data used in google earth engine cloud-based platform and count the number of fires from 2000 to 2023 for the whole country dividing it into grids and validating it by ACLED data to find some correlation between fires and conflicts. The current study main objective is developed real time monitoring system for wars like Sudan where there is little media coverage.

Another work of Hamilton and Drake (2018) explored the efficacy of the MODVOLC algorithm, which utilizes MODIS satellite data to monitor conflict-related fires during the Syrian Civil War. MODVLOC uses infrared satellite data acquired by NASA’s MODIS instrument to monitor earth’s surface for the thermal emission signature of volcanic eruptions, wildfires, and anthropogenic heat sources (e.g., gas flares) (Wright et al., 2002)

By comparing pre-war and post-war fire detections, the study highlighted a significant increase in hotspots within urban areas following the onset of conflict, as opposed to rural regions. Despite identifying a substantial number of conflict-related heat events, the study noted several limitations. MODVOLC often failed to detect fires in areas with documented violence and sometimes recorded hotspots in regions without reported incidents. Additionally, the correlation between MODVOLC detections and news coverage was weak, suggesting that some events detected by the algorithm were either not reported or misidentified as conflict related. These findings underscore the potential and challenges of using satellite-based fire detection in conflict monitoring, indicating the need for integrating satellite data with other reporting methods to improve accuracy and reliability (Hamilton and Drake, 2018)

The main weakness of the study is the failure to address how to detect conflict-related fires, especially in urban areas where tall buildings can obstruct the sensors' view which lead to underreporting of fire events in densely populated areas. Also there is an issue with false positives like the algorithm might record hotspots in areas without any reported conflict incidents, leading to potential misinterpretations of the data. This issue underscores the importance of corroborating satellite data with ground reports which leads to weak correlation between MODVOLC detections and media reports, indicating that some detected events may not be covered by international news. This constraint suggests that relying solely on news reports for validation could result in an incomplete analysis of conflict-related fires. Also, the algorithm does not distinguish between conflict-related fires and other sources of heat such as industrial activities or accidental fires. This impediment necessitates additional contextual analysis to accurately attribute detected fires to conflict activities.

Therefore, the second part of current research aims to identify different types of fires including the anomalous fires which are conflict induced by calculating the z-score value.

Tomchenko et al. (2023) paper conducted an important study on the environmental impact of the Ukraine war, specifically focusing on fires caused by military activities. Utilizing remote sensing technologies, particularly the VIIRS spectroradiometer and data from NASA's FIRMS, the study effectively monitored and assessed fires in conflict zones. The research revealed valuable insights into the spatial distribution and seasonality of fires, demonstrating the efficacy of integrating satellite imagery with traditional ground-based data for environmental monitoring. However, the study also highlighted several limitations, such as potential inaccuracies in data due to missed smaller fires or misinterpreted heat sources, limited ground verification, and a focus on regions near the front line.

Well these findings are relevant to the Sudan conflict suggesting that remote sensing can be a powerful tool for monitoring destruction caused by fires in the form of arsons and by using the dataset from ACLED and GEE cloud platform will make the research more reliable.

Conclusion

The ongoing civil war in Sudan is faced with different challenges and to investigate the damage caused by fires and number of people displaced can be done with the help of ACLED data and FIRMS fire data which uses near real time active fire locations using the standard MODIS MOD14/MYD14 fire and thermal anomalies product. Moreover a spatio-temporal study can help in knowing about changes happened in past 20 years. This research can help all the war-torn regions in world which have very less media attention and struggling with similar kind of civil war. This paper can be used to inform policy and humanitarian response in Sudan.

# Study Area and Data

The study area is country Sudan. The entire country is in war so for the analysis FIRMS fire data and ACLED conflict data is taken for the year 2000 to 2023. The most affected areas in the year 2023-2024 are Khartoum, El-Geneina, El-Fasher in Darfur, Wad Al-Nura in Al Jazirah and Nyala. These are the areas with most population.

A map of the state of south africa

Description automatically generated

Figure2 : Population density Map of Sudan

**DATA**

For this research all the dataset used are open source.

1. **FIRMS**

The study makes use of FIRMS data which can be useful when working to verify statements by actors on the ground, even in in circumstances where little or no imagery or footage is available. The FIRMS data is derived from MODIS. The present study uses the earth engine version of the FIRMS dataset which contains the NASA's Land, Atmosphere Near real-time Capability for Earth observations (LANCE) fire detection product in rasterized form. The near real-time (NRT) active fire locations are processed by LANCE using the standard MODIS MOD14/MYD14 Fire and Thermal Anomalies product. The resolution is 1000 meters. (Earth Science Data Systems, 2021) Each active fire location shows the centroid of a 1km pixel which is then captured by the algorithm containing one or more fires within the pixel. The data is rasterized in the form of 1km bounding box for each FIRMS active fire point; pixels in the MODIS sinusoidal projection that intersect the FIRMS BB are recognized; or if multiple FIRMS BBs intersect the same pixel, the one with higher confidence is retained and in case of tie the brighter one is retained. So the satellites take a picture of events as they pass over earth. Each hotspot or active fire detection indicates the center of a pixel identified as containing one or more fires or other thermal anomalies, such as volcanic activity. Therefore, the “location is the center point of pixel and not the co-ordinates of actual fire.( Earth Science Data Systems, N., 2021)

Figure3: MODIS Fire ground Observation

The shapefiles are in the geographic WGS84 projection. Some limitation of active fire pixel location is that it may not always be the most appropriate source of fire related information and the data do not provide any information on cloud cover or missing data.

1. ACLED Data (Raleigh, Kishi and Linke, 2023)

The ACLED is a project focused on the detailed collection, analysis, and crisis mapping of conflict data. It gathers information on the dates, actors, locations, fatalities, and types of all reported political violence and protest events globally. The ACLED team analyzes this data to describe, investigate, and test conflict scenarios, providing both the data and their analyses freely to the public. It is recognized as the highest-quality and most widely utilized near real-time source for political violence and protest data worldwide. The data is updated weekly, covering events up until the most recent Friday.

For the research ACLED data was downloaded for the year 2000 to 2023 as a CSV file for Sudan.

1. The Global Administrative Unit Layers (GAUL) compiles and disseminates the best available information on administrative units for all the countries in the world, providing a contribution to the standardization of the spatial dataset representing administrative units. The GAUL always maintains global layers with a unified coding system at country, first (e.g. departments), and second administrative levels (e.g. districts).

In the paper Sudan boundary has been used taken the ADM0\_Code which is the GAUL country code. (Google for Developers, 2015)

# Methods A diagram of a company Description automatically generated

Flowchart

The important point to understand about this war is that the casualties or destruction was mainly due to fires. The attacks and destruction is caused setting fire to a building or a house or burning things. Therefore, the aim of this research is develop some kind of relation between fires and conflict. Bromley, L.(2010) paper also used MODIS fire data which can be considered similar to FIRMS fire data but he did not use ACLED data which is the major difference from the previous research.

This study is using FIRMS data for calculating number of fires and ACLED data for counting number of conflict events in Sudan and then empirically a z-score is calculated which can differentiate between anomalous fires.

## Data Preparation

The objective is to prepare a long data also called panel data. The data is cleaned and validated before calculating the key components. The limitations of the data and calculation methods chosen are reflected upon and outlined in the assumptions.

The methodology in the research has been divided into two parts –

The first half of data preparation is done in Google Earth Engine(GEE) cloud computing platform. It analyzes the data that is stored remotely, each time some code is written it is send to server and speeds up the process as computation is not done locally.

While the study area is Sudan and to get the boundary of Sudan in GEE data catalogue there is Global Administrative Unit Layers (GAUL) which gives the boundary for Sudan and level 0 is picked which gives the country level boundary.

The best way to deal with the entire country is to divide it into grids so that each grid will cover data for all the years from 2000 to 2023. For defining the grid it’s important to define the scale of grid and the projection. The scale of the grid is taken (50 km x 50 km) in size. The coveringGrid() function in GEE script creates a grid that covers the entire geometry of Sudan. By dividing it into grids a long data is created which will be helpful for analysis.

A map of the state of sudan

Description automatically generated

Figure 4: Grid Formation of Sudan in GEE.

The code sets the projection to (EPSG:4326) which is a common geographical coordinate system. Firms fire data is loaded for the year 2000 to 2023 with 1000m resolution raster image and a threshold of 300 is taken that minimum range based on the value of T21 which is the brightness temperature measured in the thermal infrared band at 21µm. this band is sensitive to heat and is used to detect high temperatures, such as those found in fires.

After the grids are formed, the fire count can be calculated for every grid for each year. Now these fires can be anything like agricultural fire, forest fire, recreational fire and conflict fire. In order to get the anomalous induced fires there is a need for some outside resource to validate and match it with conflicts. In supporting this theory ACLED data is used which accounts for conflicts in the country.

The advantage is that both datasets are available starting from the year 2000, allowing us to examine historical data over the years and compare it to 2023, the conflict year, to see how it differs from past records.

## Data Cleaning

The FIRMS satellite data is used in GEE without any pre-processing but for ACLED data before using it on GEE it has been cleaned and filtered out. The original data consist of 42,156 rows and 32 columns. Due to the volume of data and the need of data manipulation, data cleaning was undertaken using Pandas 2.1.0 and NumPy 1.24.4 package. The steps undertaken were:

* Ensuring datatypes are correct,
* Removing any duplicates
* Identification and removal of missing data
* Checking the column names.
* Filtering on the basis of column **sub\_event\_types** and dropping few rows which seems non-violent or not fire related. Also filtering another column geo\_precision whose value is from 1 to 3 indicating the level of certainty of location recorded for event. Thus dropping all rows with geo\_precision = 3

A blue and white chart with text

Description automatically generated

Figure 5: ACLED Events and Sub-event types (ACLED, 2023)

**HYPOTHESIS**

The hypothesis theory in this study is that all the ACLED events should match with the anomalous fire events and then it can be conclude that there is a correlation between fires and ACLED events.

## Building the Panel Dataset

The current investigation uses these dataset to make a panel dataset where each grid for every year will give one column of fire count and other column of ACLED count. So for year 2000 grid1 will have these inputs then year 2001 for grid 1 will have similar information and so on. Every row is a time instance. Every year has a different layer from where the fire incidents can be seen in each year. Its an iterative process. Table 1 has columns grid\_id, year, fireCount, acled\_count, acled\_event\_ids and respective geometry of fire data and acled data. With this type of dataset conflict induced fires can be calculated by calculating the z-score of anomalous fires and regression can be performed to establish some kind of correlation between fire data and ACLED data.

A screenshot of a computer

Description automatically generated

TABLE 1: PANEL DATAFRAME

## Creation of Geo data frame

The above data frame consist of geographic data stored as JSON strings for ACLED and fire. To perform spatial operations and analyses on the data a Geo Data Frame is required. The above data needs to be converted into Shapely geometry objects.

A screenshot of a computer

Description automatically generated

Table 2: Geo data frame

## Initial Data Exploration

Understanding the data being used is key before undertaking further analysis. Plotting a graph for input data helps in visualization data relationships.

A graph of a graph showing the number of the company's sales

Description automatically generated with medium confidence

Figure 6: Trend of the Input Data

The above figure shows the pattern of the input data. In red is the FIRMS data and in blue is ACLED conflict data. The trend can be seen for all the years from 2000 to 2023. The long term average is also taken which gives the mean value of a dataset over a long period. Few peaks can be seen for ACLED in year 2004 and year 2023 which is exactly showing that sudden peak compared to the historical data. This shows that how the conflicts have peaked in these years.

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

Figure 7: Map of Sudan for Year 2003, 2004 and 2023

The above maps clearly shows the spread of fire data through the grids all over Sudan and ACLED events. The empty cells are the zero fire count so it’s easier to distinguish between actual events and no events.

The first two maps of 2003 and 2004 shows conflicts occurred in Darfur, Sudan which started in latter half of 2003 and continued till early 2004. Here the ACLED map clearly shows a yellow grid which is 100 above conflicts. Most of the conflicts are concentrated in the west Darfur region.

While year 2023 which is the current conflict year, a peak can be seen in red in figure 5 and that can be seen on the map also specially one grid is all yellow which is the capital city Khartoum which has more than 4000 ACLED events. So, all the conflicts are concentrated there. Therefore, a geographical shift can be seen in the history of Sudan conflict from west to all over Sudan especially concentrated in the capital of Sudan. It looks like the fighting has further moved North and is centered around Khartoum.

### Log Transformation

Both ACLED and Fire counts have asymmetric distribution with extreme right skew. Log transformation (using the natural logarithm) was applied to reduce skewness and stabilize variance. Post log transformation, the data is still skewed however with slightly shorter tail. There are signs of heteroskedasticity in the ACLED count (response variable), specifically as variance is not constant for different values of Fire count.

|  |  |
| --- | --- |
| A graph of a graph with red dots  Description automatically generated | A graph of a fire count  Description automatically generated |

Figure 8: Scatter Plot (Before Log and After Log)

## Anomalous Fires: Estimating optimum Z-score

The second part of the methodology is distinguishing between the anomalous fires. The anomalous fires are those fires which happened due to certain extreme events like in this case is conflicts. These anomalous fires will be conflict induced fires. The panel data is divided into two parts from year 2000-2022 that is the pre-conflict period separating it from the current year that is 2023 the conflict years. The shape of the data from 2000-2022 has 19550 rows and year 2023 has 850 rows.

These are the following steps taken

Step 1: Created sub-dataset for historical period (2000-2022) and current conflict year (2023)

Step 2: Binarize both datasets (each grid and all years) based on ACLED-count (1 for fighting and 0 for no fighting). Store the binary values in acled\_binary column.

Step 3: Calculate the mean and std. deviation for 2002-2022 period for each panel. Took overall mean and std (acled\_conflict 0 and 1).

Step 4: Calculate z-scores for 2023

Applied percentile method to arrive at optimum z-score, which was validated using confusion matrix and calculating accuracy, precision and recall values. Therefore with this calculation of z-score all the values above the threshold value of 1.48 can be considered as conflict induced fires. So the actual fire count known and the z-score known conflict induced fires can be calculated with these equation.

The kernel density estimation (kde) plot, visualizes the distribution of historical data from 2000 to 2022 that’s the pre-conflict period and year 2023 which is the conflict year. From this plot an empirical value of z-score can be selected and validated by binarizing the ACLED data with ‘0’ no fighting and ‘1’ fighting.

The MODVOC algorithm used in (Hamilton and Drake, 2018) paper does not distinguish between fires which are conflict induced and other fires such as industrial activity fires or accidental fires. Therefore calculating the z-score and choosing threshold can solve this problem and count of conflict fires can be taken out.

## Panel Regression

The final dataset for 20400 rows. The maximum fire count is 44 and the ACLED count is 3899. Map of ACLED for the year 2023 in Figure 7 a yellow grid can be spotted which is Khartoum showing the highest number of conflicts happening. As Khartoum is the capital city of Sudan and most populated so all conflicts are concentrated near that area and overall the population is not evenly distributed. Therefore, for further analysis it’s important to take log of the fire-count and ACLED-count.

After taking log, a Pooled OLS regression model is a good starting point and can act as a reference model. Our dependent (endogenous) and explanatory (exogenous) variables are as follows:

Dependent variable y = count of ACLED events in grids

Explanatory variable x = count of FIRMS fires in grids

Step 1: Transform data into right format for use in regression model including multi index data, with grid\_id (entity) and year (period) index.

Step 2: Apply regression models: PooledOLS, PanelOLS (Fixed-effects) and Random Effect model.

Step 3: Check goodness of fit – normality, heteroskedasticity, correlation etc.

* **Pooled OLS**: This model assumes that there are no individual-specific effects or time-specific effects. It treats the data as if it were a single large cross-sectional dataset, ignoring any panel structure.
* **Fixed Effects (Entity and Time)**: This model controls for unobserved heterogeneity by allowing for individual-specific (entity) and time-specific fixed effects. It captures the effects that vary across entities and over time but are constant within an entity or time period.
* **Random Effects**: This model assumes that the individual-specific effects are random and uncorrelated with the independent variables. It allows for variation across entities but assumes that these variations are randomly distributed.

The pooled OLS regression model equations as follows:

The general equation is like this

A black and white rectangular sign with a letter x

Description automatically generated

**Fixed effect Model**

The Fixed Effects regression model is used to estimate the effect of intrinsic characteristics of individuals in a panel data set. Examples of such intrinsic characteristics are genetics, acumen and cultural factors. Such factors are not directly observable or measurable but one needs to find a way to estimate their effects since leaving them out leads to a sub-optimally trained regression model. The Fixed Effects model is designed to address this problem. (Date, 2022)

The general equation is

A black and white rectangular sign with letters and numbers

Description automatically generated

The above equation suggests an approach for constructing the following two kinds of models — the Fixed Effects model, and the Random Effects model depending on whether or not the Covariance term in the above equation is zero, i.e. whether or not the unobservable effects are correlated with the regression variables.

In this model, it is assumed that the unobservable individual effects are correlated with the regression variables. In effect, it means that the Covariance and in the above equation is non-zero. It is also assumed that the bias introduced due to the omission of the unit-specific factors is group-specific.

To compensate for this bias, a group-specific intercept is introduced called into the model. is assumed to act in a direction that is opposite (in a vector sense) to the effect of the omitted-variable bias.

A black and white sign with a plus and a symbol

Description automatically generated

Here is assumed constant across all time periods.

# ‌Results

## Empirical selection of z-scores and validation.

A graph of a graph

Description automatically generated

Figure 9: KDE Graph of Z-score

The graph shows the kernel density estimation (KDE) of z-scores across different time periods and conditions: historical data for pre-conflict period (2000-2022), and latest conflict data (2023) with conflict and no-conflict grids.

* Red Curve: Represents the z-score density distribution in pre-conflict the historical period (2000-22)
* Blue Curve: Represents the z-score density for no-conflict grids in 2023
* Purple Curve: Represents the z-score density for conflict grids in 2023

Both the historical (red) and current no-conflict (blue) distributions peak around a z-score of 0. This indicates that, historically and currently, most no-conflict grids have z-scores close to the mean. The historical pre-conflict distribution (red) has a higher and narrower peak, suggesting a more concentrated distribution around the mean with less variance. The current no-conflict distribution (blue) is slightly broader, indicating that in 2023, there is more variation in the z-scores for no-conflict areas compared to the historical data. The similarity between the historical no-conflict (red) and current no-conflict (blue) distributions suggests that the z-score distribution in no-conflict areas has not changed drastically from the historical norm.

The current no-conflict (blue) and conflict (purple) distributions are somewhat similar, but the conflict distribution is slightly broader and more skewed towards higher z-scores. This suggests that, in 2023, conflict areas tend to have higher z-scores compared to no-conflict areas, indicating more extreme conditions or outlier values in the conflict regions. The conflict distribution (purple) extends further to the right, indicating that some conflict grids have significantly higher z-scores compared to no-conflict grids. This is consistent with the idea that conflict areas might be associated with more extreme or unusual events.

The broader distribution of z-scores in 2023, particularly in conflict areas, indicates that z-scores are picking up more variability in conflict regions, which might be useful for identifying or predicting conflict-prone areas. The right tail of the purple curve (conflict) is further used to estimate z-score threshold to distinguish conflict and no-conflict grids.

Therefore we could empirically select a threshold where the conflict and no-conflict distributions diverge the most, potentially around the higher z-scores where conflict is more likely.

**Using 95th percentile z-score on historical data, the z-score threshold is estimated is 1.48253.**

Using the estimated z-square threshold, predicted outcomes were compared with the actual outcomes in the current data. This can be done using a confusion matrix or other evaluation metrics.

|  |  |
| --- | --- |
|  |  |
| Table 3: Confusion Matrix |  |

At z-score threshold (=1.48253), this classification model is highly accurate when predicting no conflict, with a high precision, recall, and F1-score for this class. However, it struggles significantly with identifying conflicts. The precision, recall, and F1-score for class `1` are all quite low. This indicates that the model often fails to correctly identify conflicts, and when it does predict a conflict, it is often incorrect.

The overall accuracy of 78.7% may seem high, but this is largely driven by the model's ability to correctly identify non-conflict grids. The poor performance in detecting conflicts (class `1`) suggests that the model is not reliable for this class, potentially due to class imbalance or other model limitations.

## Panel Regression

A comparison of five different regression models that explain the log-transformed count of a dependent variable (log\_acled\_count) using the independent variable log\_fireCount. All models use the same dataset with 20,400 observations. Pooled OLS used a clustered covariance estimator, at the entity level, to account for potential within-entity correlation. The Fixed Effect and Random Effects models used unadjusted covariance estimators.

A screenshot of a computer

Description automatically generated

Table 4: Model Comparision

**R-squared:** The R-squared results indicate that the observed independent variable explains very little of the variation in the dependent variable. The baseline Pooled OLS indicates that approximately 3.31% of variation in Log (ACLED count + 1) is explained by Log (Fire count + 1).

In the Fixed effect model (entity and time control), the variation explained by Log (Fire count + 1) was only 0.18% (very low), whereas for Fixed effect model (time control) the variation explained was comparatively better at 3.23%, indicating that fixed effects across time periods have better explanatory power in comparison to fixed effects across entities. The Random effect model also explained a low variance of dependent variable. In other words, the unobserved independent variables would potentially explain more variation in dependent variables.

**F-statistic and P-value:** All 5 models have significant F-statistic (p-value = 0.000), indicating that the models as whole are statistically significant. The F-statistic is highest in the Pooled OLS (F-statistic = 697.5) and Fixed Effect\_T (F-statistic = 679.8) models, suggesting these models provide the most statistically powerful explanations among those tested.

**Coefficient estimates and t-statistics:** The intercept is positive and significant across all models. The magnitude of the constant varies slightly across models, with the Fixed Effect (entity and time control) model having the highest value of 0.0937 with highest t-statistic of 20.261.

**Log (Fire count +1) (Independent Variable):** The coefficient is positive and statistically significant in all models, indicating that higher fire counts are associated with higher values of ACLED counts. Though the magnitude of this effect varies by models, the coefficient establishes a mild relationship between both dependent and independent variables.

Pooled OLS: 0.0714 (most significant)

Fixed Effect (entity and time control): 0.0333 (lowest)

Fixed Effect (entity control): 0.0483

Fixed Effect (time control): 0.0697

Random Effects: 0.0551

To further establish the relationship, various test are used to check goodness of fit of each model. The results of goodness of fit test are in the table.

**Residual Errors:** All models have mean residual errors that are extremely close to zero, indicating that they are **all models are nearly unbiased in their predictions**.

**Normality:** The comparison Q-Q plots of the residuals provide visual evidence that all models deviates from normal distribution, with data at the lower end (negative values) fitting the theoretical quantiles of a normal distribution, as indicated by points lying close to the red diagonal line. There is a significant upward deviation at the upper end of the quantiles, suggesting that the data has heavy tails or extreme values (outliers) compared to a normal distribution. Visually, the distribution of **Fixed Effect (time control) appears similar to Pooled OLS.**

The Jarque-Bera Test Statistic was used to quantitatively validate the deviation from the normal distribution. The test statistics for models range from 400,000 to 857,000, which suggest significant greater deviation from the normal distribution, with higher value indicating greater deviation from normality. As all p-values are 0.0, the null hypothesis (which states that the data is normally distributed) can be rejected with a very high level of confidence. However, test statistic suggests that **Fixed Effect (entity and time control)** perform better compared to rest of the models.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Mean Value of Residual Errors** | **Q-Q Plot** | **Jarque-Bera Test Statistic** | **Raw Residuals Vs x** | **White Test** | **Raw Residuals Vs y** | **Pearson's r test of correlation**  **(Residual errors vs y)** |
| **Pooled OLS** | -1.88e-17 |  | Test stat: 856862.11  p-value: 0.0 |  | LM-Stat: 460.511  LM p-val: 1.002e-100 |  | Pearson's r: 0.9833  p-value:: 0.0 |
| **Fixed Effect (entity and time control)** | 3.99e-17 |  | Test stat: 399992.55  p-value: 0.0 |  | LM-Stat: 576.690  LM p-val: 5.934e-126 |  | Pearson's r: 0.6981  p-value:: 0.0 |
| **Fixed Effect (entity control)** | 6.41e-17 |  | Test stat: 451395.59  p-value: 0.0 |  | LM-Stat: 614.268  LM p-val: 4.105e-134 |  | Pearson's r: 0.7182  p-value:: 0.0 |
| **Fixed Effect (time control)** | 2.51e-17 |  | Test stat: 804333.13  p-value: 0.0 |  | LM-Stat: 445.300  LM p-val: 2.015e-97 |  | Pearson's r: 0.9692  p-value:: 0.0 |
| **Random Effects** | 3.48e-18 |  | Test stat: 596436.03  p-value: 0.0 |  | LM-Stat: 570.068  LM p-val: 1.630e-124 |  | Pearson's r: 0.8420  p-value:: 0.0 |

**Heteroskedasticity:** White test was used to detect heteroscedasticity in the all models. Heteroscedasticity occurs when the variance of the errors (residuals) in a regression model is not constant, which can lead to inefficient estimates and invalid inference (e.g., confidence intervals, hypothesis tests). The LM-Stat values range from approximately 445 to 614, are relatively large, suggesting that there may be significant heteroscedasticity in the models being tested. A higher LM-statistic indicates a stronger presence of heteroscedasticity in the model. Comparatively, the LM-statistic is lowest for **Fixed Effect (entity and time control) and Fixed Effect (entity control)** models, indicating relatively lesser heteroskedasticity. The extremely low p-values (effective 0) indicate that the null hypothesis of homoscedasticity can be rejected with a very high level of confidence. The results are visually confirmed by plots of residual against the independent variable (x), with better convergence of residuals with increasing x in **Fixed Effect (entity and time control) and Fixed Effect (entity control)** models.

**Correlation:** A high Pearson’s r value close to 1 indicate a strong positive linear relationship between the residual errors and dependent variable (y), indicating the models are misspecified, potentially missing key variables could be missing or possibly due to non-linearity or omitted variable bias. The **Fixed Effect (entity and time control)** has the lowest Pearson’s r = 0.6981 (still significantly correlation), but suggest the model is better equipped at identifying the influence of unobserved independent variables across entity and time period. The p-values (0.0 for all models) indicate that the correlations are statistically significant. The results are visually confirmed by plots of residual against the dependent variable (y), with residuals more spread out with increasing y in **Fixed Effect (entity and time control) and Fixed Effect (entity control)** models.

# Discussion

This study was set out to answer the following questions -

* How to identify conflict induced fires in Sudan by using FIRMS data and ACLED conflict data?
* Compare panel data regression models and suggest best modelling approach for prediction of conflicts using FIRMS data.

This study has demonstrated the use of remotely sensed FIRMS fire data and ground truth ACLED conflict to efficiently identify conflict induced fires in conflict zones of Sudan. The study used Google earth engine to create a geodata set in the form of a panel dataset covering the entire geography of Sudan in 20400 grids (50 km x 50 km each) with associated information on fires and conflicts occurring between 2000 and 2023 (24 years).

* Global coverage
* Coverage vast area coverage
* Remote areas
* Media restrictions

and Google earth engine can be used effectively to monitor fire outbreaks in conflict zones like Sudan. It also discussed about how combining these two dataset one raster dataset FIRMS and the other vector ACLED can be used to identify conflict induced fires from the regular fires and the relationship between fires and conflict. The results clearly showed that by calculating a threshold value of z-score. An empirical value is selected based on the higher side of threshold whether fires are conflict or no conflicts.

Therefore, with FIRMS fire data and ACLED conflict data a real-time monitoring system which can be build which can work similar to Sondre’s Ukraine Fire Model which uses statistical techniques from machine learning and publicly available data on temperature anomalies to detect war events.

Some limitations of the study is that the confusion matrix calculated for the classification model did show class imbalance where class ‘1’ (conflict) is underrepresented. To improve the performance techniques like oversampling, under sampling, or using different evaluation metrics (like ROC-AUC) could help. Also tuning the model or trying different algorithms might perform better in detecting the minority class(‘1’). Another thing which can be done is that synthetic data can be generated data using methods like SMOTE (Synthetic Minority Over-sampling Technique) to create more examples of conflict areas, enhancing the model's exposure to the patterns associated with conflict.

As the threshold value of 1.48253 is based on the 95th percentile of historical data. It can be further optimized using ROC curve analysis, precision-recall trade off, tail distribution analysis.

The current model provides a good foundation but needs significant refinement, especially in accurately predicting conflict areas. By addressing class imbalance, optimizing the z-score threshold, incorporating advanced modeling techniques, and continually validating the model, you can enhance its reliability and accuracy in predicting conflict-prone areas.

The Fixed Effect model, despite its low R-squared, is valuable because it provides insight into the relationship between `log\_fireCount` and `log\_acled\_count` while controlling for unobserved, time-invariant factors. The statistical significance of the `log\_fireCount` coefficient indicates that the model is capturing a meaningful relationship, making it a crucial part of the analysis.

### \*\*2. Refine the Z-Score Threshold:\*\*

- \*\*Threshold Optimization\*\*: The threshold of 1.48253 is based on the 95th percentile of historical data. You could explore optimizing this threshold using different approaches like:

- \*\*ROC Curve Analysis\*\*: Generate a Receiver Operating Characteristic (ROC) curve to determine the optimal z-score threshold that maximizes the true positive rate (recall) while minimizing the false positive rate.

- \*\*Precision-Recall Trade-off\*\*: Given that conflict prediction is critical, consider adjusting the threshold to improve the precision and recall for conflict areas, even if it slightly reduces accuracy in predicting no-conflict areas.

- \*\*Tail Distribution Analysis\*\*: Given that conflict regions tend to have higher z-scores, further analyze the tail distributions to fine-tune the threshold for better conflict identification.

- \*\*Addressing Class Imbalance:\*\* If class `1` (conflict) is underrepresented, techniques like oversampling, undersampling, or using different evaluation metrics (like ROC-AUC) could help improve performance.

- \*\*Model Tuning:\*\* Consider tuning the model or trying different algorithms that might perform better in detecting the minority class (`1`).

Given the detailed analysis and the identified challenges, here are recommendations for improving the model and further actions to take:

### \*\*1. Address Class Imbalance:\*\*

- \*\*Resampling Techniques\*\*: Since the model struggles with identifying conflict areas (likely due to class imbalance), you can use techniques like oversampling the minority class (conflict grids) or undersampling the majority class (no-conflict grids). This could help the model learn better from the conflict data and improve its ability to predict this class.

- \*\*Synthetic Data Generation\*\*: Consider generating synthetic data using methods like SMOTE (Synthetic Minority Over-sampling Technique) to create more examples of conflict areas, enhancing the model's exposure to the patterns associated with conflict.

### \*\*3. Enhance Model Complexity:\*\*

- \*\*Feature Engineering\*\*: Incorporate additional features that might capture underlying patterns associated with conflict areas. For example, spatial features (proximity to conflict zones), temporal trends, or even interaction terms between existing variables.

- \*\*Advanced Modeling Techniques\*\*:

- \*\*Ensemble Methods\*\*: Explore ensemble methods like Random Forests, Gradient Boosting Machines (GBM), or XGBoost, which might capture non-linear relationships and interactions better than a simple threshold-based model.

- \*\*Deep Learning Models\*\*: If the dataset is large enough, deep learning approaches such as Convolutional Neural Networks (CNNs) for spatial data or Recurrent Neural Networks (RNNs) for temporal sequences might capture more complex patterns.

### \*\*4. Model Evaluation and Validation:\*\*

- \*\*Confusion Matrix and Metrics\*\*: Regularly use a confusion matrix to evaluate the model's performance, focusing on precision, recall, and F1-score for conflict grids. Also, consider using metrics like the Area Under the Precision-Recall Curve (PR AUC), especially in cases of class imbalance.

- \*\*Cross-Validation\*\*: Implement k-fold cross-validation to ensure that the model generalizes well to unseen data, and to prevent overfitting to the specific training set.

### \*\*5. Investigate and Incorporate Additional Data:\*\*

- \*\*External Data Sources\*\*: Integrate external data that might have predictive power for conflicts, such as socioeconomic indicators, political instability metrics, or satellite imagery capturing environmental changes (e.g., drought, deforestation).

- \*\*Temporal Analysis\*\*: Conduct a time-series analysis to understand how the z-scores and conflict likelihood evolve over time, which might offer predictive insights.

### \*\*6. Monitor and Iterate:\*\*

- \*\*Continuous Monitoring\*\*: As new data becomes available, regularly update the model and the z-score threshold to reflect any changes in the underlying patterns of conflict and no-conflict areas.

- \*\*Model Feedback Loop\*\*: Incorporate a feedback loop where the model's predictions are continuously compared against actual outcomes, and the model is retrained or adjusted based on performance.

### \*\*7. Consider Non-Statistical Factors:\*\*

- \*\*Expert Input\*\*: Incorporate domain expertise into the model development process. Experts in conflict studies might offer insights into factors that aren't captured by purely statistical approaches, which could be integrated as additional features or adjustment factors.

- \*\*Scenario Analysis\*\*: Use the model to conduct scenario analyses, testing how changes in certain variables (e.g., increasing fire counts or other stressors) might influence the likelihood of conflict, and adjust your strategy accordingly.

### \*\*Conclusion:\*\*

The current model provides a good foundation but needs significant refinement, especially in accurately predicting conflict areas. By addressing class imbalance, optimizing the z-score threshold, incorporating advanced modeling techniques, and continually validating the model, you can enhance its reliability and accuracy in predicting conflict-prone areas.

**Recommendation:**

Given the performance of the models across various criteria, here’s a recommended approach:

1. \*\*Use a **Robust Fixed Effect Model** with **Entity and Time Controls**:
   * **Why**: Despite its low R-squared, this model shows the lowest Pearson’s rrr and performs relatively better in terms of heteroskedasticity and normality. It accounts for unobserved heterogeneity across entities and time, which might be crucial given the data structure.
2. **Address Model Misspecification**:
   * **Non-linearity**: Consider exploring non-linear models or transformations of the independent variable to capture the relationship more effectively.
   * **Omitted Variables**: Investigate potential omitted variables that might better explain the variation in the dependent variable.
3. **Correct for Heteroskedasticity**:
   * Use robust standard errors or consider models like Generalized Least Squares (GLS) to account for the heteroskedasticity detected by the White test.
4. **Alternative Approaches**:
   * **Mixed-Effects Models**: Given the data structure (entities over time), consider using mixed-effects models that can handle both fixed and random effects, providing more flexibility in capturing unobserved heterogeneity.
   * **Machine Learning Models**: If the goal is prediction rather than inference, machine learning models like Random Forest or Gradient Boosting could be explored to capture non-linear relationships and interactions.

**Conclusion:**

While no model performs perfectly, the **Fixed Effect Model with Entity and Time Controls** appears to be the best starting point, given its better handling of residual correlations and slightly improved heteroskedasticity. However, further refinement and testing of alternative models are recommended to address the identified issues and improve model performance.

In summary, while the model is effective at predicting no conflicts, it performs poorly in detecting actual conflict instances, which is a significant concern depending on the importance of accurately identifying conflicts in your application.

Yes, the results do provide valuable insights about the Fixed Effect model, even though the R-squared value is low. Here's how to interpret the significance and usefulness of the Fixed Effect model in this context:

### 1. \*\*Understanding the Low R-Squared Value:\*\*

- \*\*R-Squared Interpretation\*\*: The R-squared value, particularly the "Within R-squared" for Fixed Effects, measures the proportion of the variance in the dependent variable (within entities over time) that is explained by the independent variables. A low R-squared does not necessarily mean that the model is ineffective; it simply indicates that the independent variables explain only a small portion of the variation in the dependent variable within entities.

- \*\*Nature of the Data\*\*: In panel data, especially with many entities and over time, it’s common for the within R-squared to be lower because individual-level variations may be more influenced by factors not captured by the model, such as unobserved heterogeneity.

### 2. \*\*The Role of the Fixed Effect Model Despite Low R-Squared:\*\*

- \*\*Controlling for Unobserved Heterogeneity\*\*: The Fixed Effect model's primary strength is its ability to control for time-invariant characteristics of the entities (e.g., countries, companies, etc.) that might affect the dependent variable. This means that the Fixed Effect model provides more reliable estimates of the relationship between the independent variable (`log\_fireCount`) and the dependent variable (`log\_acled\_count`) by isolating the effect of `log\_fireCount` from these unobserved, constant factors.

- \*\*Statistical Significance of Coefficients\*\*: Despite the low R-squared, the coefficient for `log\_fireCount` (0.0267) is statistically significant with a p-value far below 0.05 (indicated by the t-statistic of 3.5873). This suggests that even after controlling for unobserved, entity-specific factors, there is a significant relationship between `log\_fireCount` and `log\_acled\_count`.

### 3. \*\*Implications for Research:\*\*

- \*\*Focus on Coefficient Significance\*\*: In Fixed Effect models, the emphasis is often on the significance and magnitude of the coefficients rather than the R-squared value. The significance of the `log\_fireCount` coefficient indicates that changes in this variable are indeed associated with changes in the `log\_acled\_count`, even when accounting for entity-specific effects.

- \*\*Utility of Fixed Effects\*\*: The Fixed Effect model is particularly useful when the goal is to understand how changes in an independent variable within an entity affect the dependent variable over time, rather than explaining variance across entities. The model's ability to control for unobserved, time-invariant factors is its primary advantage.

### 4. \*\*Choosing Between Models:\*\*

- \*\*Trade-off\*\*: While the Fixed Effect model might not explain as much of the overall variance in the dependent variable (as reflected by the lower R-squared), it provides a more accurate estimate of the relationship between the variables of interest by removing the bias from unobserved heterogeneity.

- \*\*Model Suitability\*\*: The choice of model should be guided by the research question. If the research is concerned with the effects of variables within entities over time, the Fixed Effect model is usually preferred despite the low R-squared.

### Conclusion:

The Fixed Effect model, despite its low R-squared, is valuable because it provides insight into the relationship between `log\_fireCount` and `log\_acled\_count` while controlling for unobserved, time-invariant factors. The statistical significance of the `log\_fireCount` coefficient indicates that the model is capturing a meaningful relationship, making it a crucial part of the analysis.

# Conclusion

It is hoped that this work will serve as a foundation for future investigations into how to deal with FIRMS fire data and how that can be linked to conflict datasets like ACLED. This study provides ideas on how to distinguish between conflict fires. Future work should also more comprehensively investigate different types of conflict dataset like UNOSAT and others.

This kind of study also help to understand how to improve such datasets maintaining high data quality.

A screenshot of a computer game

Description automatically generated

‌

# References

Abdelaziz, K. and Lewis, A. (2023). Sudan factions agree to extend ceasefire deal amid clashes. Reuters. [online] 30 May. Available at: <https://www.reuters.com/world/africa/heavy-clashes-sudans-capital-truce-set-expire-2023-05-29/>.

ADF (2024). Smuggled Gold Fuels War in Sudan, U.N. Says. [online] Africa Defense Forum. Available at: <https://adf-magazine.com/2024/02/smuggled-gold-fuels-war-in-sudan-u-n-says/>.

Ali, T. (2024). Sudan in Crisis: Mapping the War’s Toll on People. [online] ArcGIS StoryMaps. Available at: https://storymaps.arcgis.com/stories/ee9c6513732344448bd6337fdadc11c6 [Accessed 10 Jun. 2024].

CIR (2024). More than 200 villages and towns damaged or destroyed by fire since the start of the war in Sudan, with April the worst month on record. [online] Centre for Informati. Available at: <https://www.info-res.org/post/more-than-200-villages-and-towns-damaged-or-destroyed-by-fire-since-the-start-of-the-war-in-sudan-w>.

Date, S. (2022). Understanding the Fixed Effects Regression Model. [online] Medium. Available at: https://towardsdatascience.com/understanding-the-fixed-effects-regression-model-d2fccc2cc27e.

Doxsee, C. (2023). How Does the Conflict in Sudan Affect Russia and the Wagner Group? www.csis.org. [online] Available at: <https://www.csis.org/analysis/how-does-conflict-sudan-affect-russia-and-wagner-group>.

Earth Science Data Systems, N. (2021). MCD14DL-NRT | Earthdata. [online] www.earthdata.nasa.gov. Available at: <https://www.earthdata.nasa.gov/learn/find-data/near-real-time/firms/mcd14dl-nrt>.

Earth Science Data Systems, N. (2021). FIRMS Frequently Asked Questions | Earthdata. [online] www.earthdata.nasa.gov. Available at: https://www.earthdata.nasa.gov/faq/firms-faq#ed-wfm-download [Accessed 31 Jul. 2024].

The Economist and Solstad, Sondre (corresponding author), 2023. The Economist war-fire model. First published in the article "A hail of destruction", The Economist, February 25th issue, 2023.

Firms CSV- firms.modaps.eosdis.nasa.gov. (n.d.). NASA-FIRMS. [online] Available at: <https://firms.modaps.eosdis.nasa.gov/country/>.

Gallopin, J.-B. (2024). ‘The Massalit Will Not Come Home’. Human Rights Watch. [online] Available at: <https://www.hrw.org/report/2024/05/09/massalit-will-not-come-home/ethnic-cleansing-and-crimes-against-humanity-el>.

Google for Developers. (n.d.). FIRMS: Fire Information for Resource Management System | Earth Engine Data Catalog. [online] Available at: https://developers.google.com/earth-engine/datasets/catalog/FIRMS#bands [Accessed 31 Jul. 2024].

Google for Developers. (2015). FAO GAUL: Global Administrative Unit Layers 2015, Country Boundaries  |  Earth Engine Data Catalog  |  Google for Developers. [online] Available at: https://developers.google.com/earth-engine/datasets/catalog/FAO\_GAUL\_2015\_level0#terms-of-use [Accessed 9 Aug. 2024].

Kristof, N. (2024). Opinion | From the Embers of an Old Genocide, a New One May Be Emerging. The New York Times. [online] 15 May. Available at: <https://www.nytimes.com/2024/05/15/opinion/darfur-sudan-genocide.html>.

OCHA (2024). Sudan. [online] reports.unocha.org. Available at: <https://reports.unocha.org/en/country/sudan/>.

‌ Rickett, O. (2024). How the UAE kept the Sudan war raging. [online] Middle East Eye. Available at: <https://www.middleeasteye.net/news/sudan-uae-war-arms-trade-rsf>.

‌

Raleigh, C., Kishi, R. & Linke, A. Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. Humanit Soc Sci Commun 10, 74 (2023). <https://doi.org/10.1057/s41599-023-01559-4>

Trad, T.K. (2023). Sudan’s gold: The precious metal used to fund conflicts. [online] https://www.newarab.com/. Available at: <https://www.newarab.com/analysis/sudans-gold-precious-metal-used-fund-conflicts>.

UNHCR (2024). Sudan crisis explained. [online] www.unrefugees.org. Available at: <https://www.unrefugees.org/news/sudan-crisis-explained/>.

Wim Zwijnenburg and Ballinger, O. (2023). Leveraging emerging technologies to enable environmental monitoring and accountability in conflict zones. International Review of the Red Cross, 105(924), pp.1497–1521. doi:https://doi.org/10.1017/s1816383123000383.

‌

‌

‌Appendix:

A blue sky with stars

Description automatically generated

Population Map of Sudan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BANDS | | | | |
| Name | Units | Min | Max | Description |
| T21 | K | 300\* | 500.29\* | The brightness temperature of a fire pixel using MODIS channels 21/22. |
| Confidence | % | 0 | 100 | A detection confidence intended to help users gauge the quality of individual active fire pixels. The confidence estimate ranges between 0% and 100% for all fire pixels within the fire mask. |
| Line\_number |  | 1\* | 35302\* | Line number in the FIRMS CSV file that the pixel came from. |

Table 1: BANDS

A close up of a blue and white box

Description automatically generated