Scientific African 20 (2023) e01656

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

A fusion of structured and unstructured datasets in curating fire damage

Daniel O-Sullivan Hewlett, Moreblessings Shoko*, Brighton Chamunorwa

Geomatics Division, University of Cape Town, South Africa

ARTICLE INFO

Article history:
Received 27 November 2022
Revised 16 February 2023
Accepted 22 March 2023

Editor: DR B Gyampoh

Key words:
Unstructured data
Sustainable cities
Curation
UAV
Data fusion
Wildfires

ABSTRACT

Accurate documentation and mitigation of wildfires is significant in addressing Sustainable Goal eleven (11) as well as the African Unions Agenda 2063 Goal one (1) which resonate an intention to preserve sustainable environments, communities, and cities for future generations. Poor and unscientific documentation of wildfire leads to error prone and quickly outdated event curation, making it difficult to create comprehensive overviews of wildfire events. Although several geospatial studies have recorded advances in wildfire data curation, many of these rely on structured spatial datasets which enjoy numerous advantages but can be prone to temporal or time-voids in data. Voids in data can be mitigated through interpolation techniques, but these are not without challenges for the modellers tasked with optimally filling the voids. On this backdrop, this paper therefore aims to test and extend strategies of wildfire reporting by investigating the strengths of fusing mainstream spatial data (including high resolution unmanned aerial vehicle (UAV) footage) with unstructured Twitter (tweets) data in order to disrupt common practise and contribute to the global agenda. The novelty of this contribution lies in its unusual hybrid data approach used to statistically establishes long- and short-term pre-fire conditions tested on the African landscape. It proceeds to illustrate how the social media data can create fascinating and yet accurate visual timelines of the events that surprisingly compared strongly to the results from structured sources. Finally, the study detects the extent of damage from the wildfire using supervised classification, burn indices and three-dimensional (3D) reconstruction. A 95% positive detection rate was reported, and it affirmed the place of unstructured data in mainstream scientific approaches such as wildfire documentation. The work presented therefore contributes towards meeting African Union's Agenda 2063 and the United Nations Sustainable Development Goals (2030) in curation, documentation and mitigation of sustainable community and environments.

Introduction

The eleventh (11th) Goal of the United Nations Sustainable Development Goals (SDG30) and Aspirations (Goal 1.7; Goal 2.2) of the African Union Agenda 2063 reflect on sustainable governance and preservation of global and regional cities and

* Corresponding author.
E-mail address: m.shoko@uct.ac.za (M. Shoko).

https://doi.org/10.1016/j.sciaf.2023.e01656

communities[1]. More specifically Targets 11.4 and 11.6 of SDG30, clearly pledge a global intention to safeguard the world's cultural and natural heritage for future generations and paying special attention to air quality and waste management to reduce the adverse per capita environmental impact of cities by 2030. These targets of sustainably fostering our world for future generations make the management of disasters such as wildfires, a key and current global challenge that requires decision and policy related mitigation. In much of the global academic, climate and media discourses, wildfires are characterized as a significantly increasing threat to socioecological safety. This assertion is in part valid due to the moderately strong statistical link between anthropogenically forced climate change and wildfire severity [8]; however, it is still largely false as scholars have depicted accurate modelling that argues that global burn area has decreased by as much as 14% since the nineth century [6]. It is also important to recognize most wildfires are anthropogenic in origin. A 2016 assessment by Balch et al of 30 years of wildfire records indicated that 84% of wildfires in the USA for example, were started by humans [20] and likewise 86% of responses to the questionnaires of qualitative study focusing on the causes of wildfires conducted in Zimbabwe indicated that the origin of wildfires in local regions were caused by direct human influence [21]. The African continent experiences wildfires frequently due to the prevalent, highly combustible, savanna grasslands [19]. According to reported data, Africa has had the lowest number of documented wildfire events since 1900, most likely due to a lack of reporting[2], highlighting a gap in wildfire data curation. This paper seeks to provide a framework to contribute to filling that gap with the fusion of structured and unstructured data.

Several studies in non-geospatial disciplines have started to illustrate the merits of combining structured and unstructured data for various applications. Agreeably, both structured and unstructured data can be consolidated to provide improved and additional mechanisms for disseminating information on wildfires, therefore, closing the gap of underreporting of wildfires in Africa and globally. Structured data is data that is available in a predefined format to facilitate processes such as retrieval and usage, whereas unstructured data is not organised for a specific purpose. Examples of prominent structured data in geospatial sciences include several products derived from remotely sensed data acquired by sensors mounted on various platforms [2,3,5]. Unstructured data is often captured by individuals with mobile devices and commonly distributed through social media and informal spaces [16,20]. The advantages of using traditional structured remote sensing as a tool in wildfire management have widely been recognized as it can used to observe dangerous landscapes all while consistently taking measurements at lower costs when compared to the equivalent ground survey costs. Typically, satellite borne sensors offer the greatest spatial coverage and many datasets are widely and openly accessible. However, low temporal and spatial resolution means some satellite sensors do not revisit areas frequently enough for continuous fire monitoring and often cannot image small scale fires which can potentially grow [1]. Furthermore, the lag time between the image being taken and being available to an end-user is long enough to make the usage of some satellite sensors limited in fire monitoring and detection [1]. Scholars like Leblon et al in their summary of remote sensing in wildfire management highlight two pre-wildfire conditions that can be typically monitored remotely by observing fuel type via vegetation mapping and fuel moisture via normalized difference vegetation index (NDVI) calculations [15]. However, some argue this is not robust and suggest other metrics [9,12,13]. The application of unmanned aerial vehicles (UAV) for wildfire management is a developing field. Airborne sensors are also valuable in filling the wildfire management gaps between ground-based methods and satellite imagery [1]. Where spaceborne or satellite platforms fall short in their manoeuvrability and temporal resolution, airborne platforms excel in that they can be maneuvered rapidly into any position including dangerous environments and can carry various sensors that are often mission specific unlike satellite platforms which have fixed payloads [18]. UAVs are commonly used to survey active wildfires but, in some cases, they are used to monitor vegetation biomass and for vegetation classification [7,14]. The documented incorporation of UAVs for assessments on infrastructure damage caused by wildfires are still relatively sparse. An advantage to performing a UAV-based building damage assessment is that it is non-invasive when compared to a ground survey and can scan building surfaces for faults and damage to an extremely high precision [9,11].

The inclusion of social media data in geospatial systems for detecting, monitoring, and mapping disasters (including wildfires) is somewhat of a novel practice and still relatively manual [23]. Of recent, using machine learning with sentiment analysis and toponym extraction of raw text, automation is a developing space and application. The aim of this study is to propose this interesting approach that blends structured (remotely sensed imagery and unmanned aerial imagery) and unstructured (social media data) datasets to map, model and quantify the effects of a fire using the case of the April 2021, Table Mountain fire as a study site. To meet its aim, this study harnesses the merits of this multiple datasets to create a holistic overview of the wildfire before, during and after its occurrence. The study manages to achieve this by establishing long- and short-term pre-fire conditions by statistically analysing imagery and auxiliary data. It proceeds to create a visual timeline of the wildfire event itself using unstructured social media data from Twitter. The study finally detects the spatial extent and damage of the wildfire through supervised classification, burn indices and 3D reconstruction of UAV data depicting fire scarred building infrastructure within the fire site to quantify the extent of the fire and impact to both natural environments and infrastructure. The 2018 Elandskraal Fire, Knysna report conducted by the Council for Scientific and Industrial Research (CSIR) is an ideal example of related data driven scientific analysis.

[1] https://au.int/agenda2063/aspirations.

[2] https://www.emdat.be/.

Table 1
Table showing main materials and data used in study (Source: Own Compilation).

Dataset	Data specifications
MODIS vegetation index data	from 01-01-2016 to 01-01-2021 were collected using GEE to garner a long-term visualisation of climate and vegetation dynamics in the region
Surface water storage data	of Steenbras Lower dam, accessed from the NIWIS dashboard, for assessing regional climate conditions.
Weather data	from Wind Guru and Meteoblue, online weather forecasting and data providers, was acquired at a 13km and 10km resolution. for April to assess the short-term conditions leading up to the wildfire.
Twitter data	comprising of both textual and image data from official sources and fire updates posted by individual citizens
Sentinel 2 data	is accessed through GEE for assessing post-fire damages.
UAV survey data	to obtain very high spatial resolution data for individual building assessment and 3D mapping.

Material and methods

Event background and study area

A wildfire, which was later determined to be anthropogenic in origin, began along the base of Devil's Peak, and spread across the slopes of the Table Mountain Nature Reserve in Cape Town, South Africa, from the 18th to the 20th of April. The wildfire resulted in the hospitalisation of five (5) firefighters, the evacuation of more than 4000 estimated residents from the surrounding suburbs and extensive damage to vegetation and property. This predominantly included the historical and nearby key facilities which are situated directly adjacent to the wildland-urban interface (WUI). Extensive firefighting efforts including approximately 250 firefighters and multiple water bombing helicopters were required to control the spread of the fire. The estimated cost of the damage to nearby infrastructure alone are estimated to be more than R500 million (USD $29 million).

This affected study area and region is predominately vegetated with Southern Afro temperate Forest on the south-eastern slope of Devil's Peak, Cape Vineyard Shale Fynbos on the north-eastern slope of Devil's Peak, Peninsula Granite Fynbos near Vredehoek and Peninsula Sandstone Fynbos along Table Mountain itself [3]. Additionally, Vegetation in the area have been shown to be good fuel sources [3]. Cape Town is located within the subtropical Mediterranean ("Csb") climate according to the Köppen climate classification. On average, it receives an average of 542 mm of rainfall annually, with warm summers and cold winters. In spring and summertime, a moderate to strong-south-easterly wind prevail. Cape Town also experiences Berg (mountain) winds which are dry-hot katabatic winds.

Pre-wildfire conditions

Much of the recorded and available information related to these fires comes from unconsolidated news articles making it difficult to create comprehensive overviews of these common events. At the time of the initial investigation of this report, no full public documentation nor scientific or forensic outputs such as fire statistics on the April 2021 Table Mountain fire had been published. Considering the significant public interest in the event, as seen in the spike in volume for the term "fire" Google on search trends in the week of the fire, there was research motivation to perform collect and test data curation approaches from the site. Table 1 shows the datasets used in the study.

As noted in similar wildfire investigations [9], Moderate Resolution Imaging Spectroradiometer (MODIS) datasets, as seen in Table 1, provide a variety of products applicable to wildfire analysis. In this study the normalised difference vegetation index (NDVI) product was considered the only vegetation indicator for climate conditions leading up to the wildfire. This is because the enhanced vegetation index (EVI) is better suited to high biomass regions [9], of which the study region in question is not. MODIS vegetation index data ranging over seven years from 2016-01-01 to 2022-01-01 were collected and plotted by day of each year using a cloud-based platform-Google Earth Engine (GEE). This was done to garner a longer-term visualization of vegetation dynamics in the region over the course of subsequent years.

In addition to using MODIS imagery for the assessment of regional climate conditions, surface water storage data for Steenbras Lower dam, the closest dam to the study region, was gathered from the National Integrated Water Information System (NIWIS) dashboard. Lastly, short term weather dynamics also incredibly influential in wildfire spread were collected and studied. This short-term data included hourly weather data from global forecast model, namely WindGuru[3] and Meteoblue[4] online weather data providers. The specific attributes included weather data included: wind speed, wind direction, gusts, temperature, precipitation, and relative humidity.

The data which was acquired, tabulated, and averaged out for April was used to assess the short-term conditions leading up to wildfire.

[3] https://www.windguru.cz/91.
[4] https://www.meteoblue.com/en/weather.

Fire timeline reconstruction from unstructured (Twitter) data

To construct an accurate timeline of the wildfire event in the absence of highly temporal satellite imagery, both textual and image data from Twitter data was gathered to map the spread of the fire. The value of using Twitter for wildfire monitoring is well established [20]. This acquisition was performed by performing searches on Twitter using filters based on keywords, hashtags, specific accounts, times, and location (via geocoding). Social media by nature is open and thus data on social media platforms are often not verified. Based on this reality, it was decided that mainly Tweets for the investigation would predominately be made from official Twitter accounts such as from South African National Parks (SANParks), Working on Fire, Table Mountain National Park (TMNP) and from formal news outlets where possible. These accounts are loosely termed “elite” or “opinion” based users, and the public acquires information from them. Elite users are important as they filter and interpret data before passing it on, keeping accurate information making them trustworthy for scientific curation [20]. During non-working hours however, Tweets from official sources were sparse and thus fire updates posted by individual citizens were then gathered to avoid breaks in data. It was agreed that ideal Tweets would be those whose bodies included toponym (place name), time and imagery information. A challenge faced when selecting the most suitable Tweets was differentiating between near-real time and retrospective information. Often a Tweet would contain an image that was taken at an earlier time in the day than the Tweet was posted. This was addressed by selecting Tweets that contained immediate language indicating that the Tweet was addressing a live event or contained temporal information in the body of the message such as the following examples:

> *“URGENT NOTICE. SANParks Table Mountain National Park requests that all hikers within the Newlands and Rhodes Memorial area evacuate with immediate effect.” - 11:45 Apr 18, 2021 (TNMP)*

> *“View from Mowbray now” - 21:00 Apr 18, 2021 (Individual citizen)*

From the 86 selected suitable Tweets, the spatial and temporal information (toponyms, timestamps, and images) was extracted and tabulated. Tweets and associated images on Twitter themselves do not have any accurate spatial metadata for privacy reasons. Therefore, for textual Tweets, the approximate coordinates of the area that the Tweet was referring to was recorded using Google Earth Pro and similarly the real-world coordinates of the fire contained in the posted images were extracted by geolocating the images and observing the orientation and notable features. A timeline of hourly resolution was constructed and populated with information on the wildfire spread from selected Tweets. In addition, the gathered short term weather data was plotted along the timeline to observe the changing weather conditions over the course of the fire. A timeline of the fire emerged from this exercise.

It must be noted that prior to the timeline reconstruction using the curated Twitter data, it was attempted to use the geotags in some Tweets to map those Tweets spatially. However, in a related study, it was noted that of all the data collected only 3% of Tweets were geotagged[5] . In addition, the geotag information is only a coarse location with no precise coordinate information which is unsuitable for highly accurate mapping. This highlights a limitation of using social media data in isolation. Although not being ‘remotely sensed’ in the traditional sense, unstructured social media data such as Tweets, include both textual or graphic content, and are remotely sourced by human sensors and can complement traditional remotely sensed data and can be a powerful tool in response to disaster events. Due to its conciseness, trend analysis, seamless hashtag integration and near real-time feed, Twitter is often the go-to source in breaking news and monitoring live events. It is observed that the volume and usefulness of social media data in the proximity of the disaster typically increases in-step with the temporal evolution of the disaster event which provides additional validation for the usage of social media as a data source. Thereafter, to improve the visualisation of the spatial representation of the wildfire timeline, it was decided that imposing the tweets and the general spread of the fire over time onto a 3D map created by draping Sentinel-2 red-green-blue (RGB) imagery over a 10m resolution digital elevation model (DEM) in ArcGIS Pro would be key. An accuracy assessment as employed on all the model derived results against ground truth (control points and knowledge of area) was conducted and accuracy recorded to give credibility to the model reliability.

Post-fire damage assessments

Burn scar mapping and Burn severity

In South Africa, there exist many biomes that create conducive environments for wildfires and thus occurrences of wildfires have been reported frequently at several locations countrywide [17]. To clearly and simply observe the spatial footprint and magnitude of the fire a burnt-unburnt classification map and a burn severity map were created using Sentinel-2 data which has 13 spectral bands and a 10m spatial resolution which allows for finer resolution mapping across multiple wavelengths. The Level-2A product from Sentinel-2 were gathered and filtered in GEE. Images from 21/03/2021, 20/04/2021 and 30/04/2021 images were selected to represent pre-wildfire, burning wildfire and post-wildfire conditions respectively. Due to the time difference between the pre and post wildfire images, harsh shadows from the topography needed to be masked out which was done using the hill shadow tool in GEE and replaced using a median blur of three previous images without

[5] http://dx.doi.org/10.1109/MIS.2013.126.

shadow. This was important as the shadows would have introduced errors in the burn scar mapping. A distinct advantage of working within GEE for wildfire analysis is the variety of tools and pre-processed datasets that exist all in one environment.

The burnt-unburnt classification was performed using the Random Forest supervised machine learning algorithm within GEE. Random forest classification uses the basic concept of decision trees but instead of using the output classification from one tree which is simplistic and often results in overfitting, the outputs from multiple decision trees are aggregated [16,20]. The selection of variables used in any classification is essential. Due to the nature of fire, it effects vegetation and landscapes in the visible, near infrared and shortwave infrared portions of the electromagnetic spectrum. The inclusion of spectral indices and textures as variables in classification is also particularly valuable as indices and textures typically are created to emphasize the spectral response relating to a phenomenon. Xulu et al found in their classification of burnt areas in the Southern Cape forests that the normalised burn ratio (NBR) had an importance metric of more than 90% [22]. However, NBR can be outperformed by other indices based on when the burn mapping takes place. Fornacca et al noted this in their comparison of the ability of different indices to detect burnt areas in North-western China [10]. Based on literature and several iterations of classification, the following variables were used in classification.

The study used 250 sampling points split according to 80:20 for training data and testing data respectively. Once all variables and training sites were selected, random-forest classification was performed using 100 trees and the default number of variables per split which was the square root of the number of variables. Several iterations of classification were performed using different spectral indices until a desirable accuracy was achieved. All index and texture calculations used in classification and the accuracy assessment was conducted in GEE. To map burn severity the Difference Normalised Burn Ratio (dNBR) product was then calculated by subtracting the post-wildfire NBR image from the pre-wildfire image and then applying a scaling factor to the result as seen by the dNBR formula. The resulting dNBR continuous raster was reclassified into discrete classes according to the burn severity classification scheme proposed by the United States Geological Survey (USGS). Due to the simplistic nature of the dNBR index, noise was present in the image which was filtered using a focal median low pass filter with a radius of two pixels. A high degree of spatial autocorrelation in the severity classes was present, meaning pixels next to one another tend to be in the same class, therefore applying a median filter to the dNBR severity classification is logical. The following equations were of importance to the calculations described above.[6,7]

$$\text{Normalised Burn Ratio [NBR]} = \frac{NIR - sSWIR}{NIR + sSWIR} \tag{1}$$

$$\text{Difference Normalised Burn Ratio[dNBR]} = (NBRprefire - NBRpostfire)*1000$$

where NIR is Near Infra-Red band data, sSWIR is short wave Infra-Red band data.

Lastly, to understand the dynamics between the fire and land use and land cover classes (LULC), the intersection between the burnt area, burn severity and the 2018 South African LULC layer was derived.

UAV based scene reconstruction of damaged building infrastructure

Imagery of the fire damaged buildings were acquired using a Phantom 4 Pro V2 consumer grade UAV. Several UAV survey flights were flown within a predetermined bounding area of the most affected buildings. A nadir series of flights were conducted with a 90° camera angle (looking straight down), flown in a grid pattern with a 70% image overlap. A second grid series of flights was also flown but with a 60° camera angle and a 70% image overlap. Lastly, a series of oblique orbital flights were performed which included flying the UAV in several elliptical orbits throughout the study area thus providing 360° views with a 60% camera angle. All flights were flown at an altitude of 70m above the relative terrain except for orbits of damaged buildings which were closer. The weather conditions over the flight period were variable with windy, overcast conditions on the 06/10/2021 and calm, clear conditions on the 07/10/2021.

The drone survey imagery further required the acquisition of ground control points (GCPs) to georeference the imagery to a high precision when performing mapping and structure from motion (SfM) photogrammetry. Using a Trimble RTK (Real-Time Kinematic) positioning system composed of a differential global navigation satellite system (GNSS) receiver in GPS (Global Positioning Systems) mode, with base mounted on a known position and a mobile rover, 22 Real Time Kinematic (RTK) ground control point (GCP) locations were recorded.

Finally, to perform accurate measurements of burnt buildings and areas using the acquired drone data, structure-from-motion (SfM) photogrammetry was performed using the open-source software package Open Drone Map (ODM) and its web interface allowed for the creation of 3D models, orthophotos and point clouds from overlapping drone imagery and the collected GCPs. ODM reconstructions were run on a local Ubuntu 20.04 system with 16 giga bytes (GB) of random Access Memory (RAM), an i7 – 6700 3.40GHz Computer Processing Unit (CPU) and an NVIDIA GTX 1080 graphics card. Area measurements of the scars on damaged buildings were computed using the 3D viewer of the ODM web interface.

[6] Source [16,20].
[7] Source [16,20].

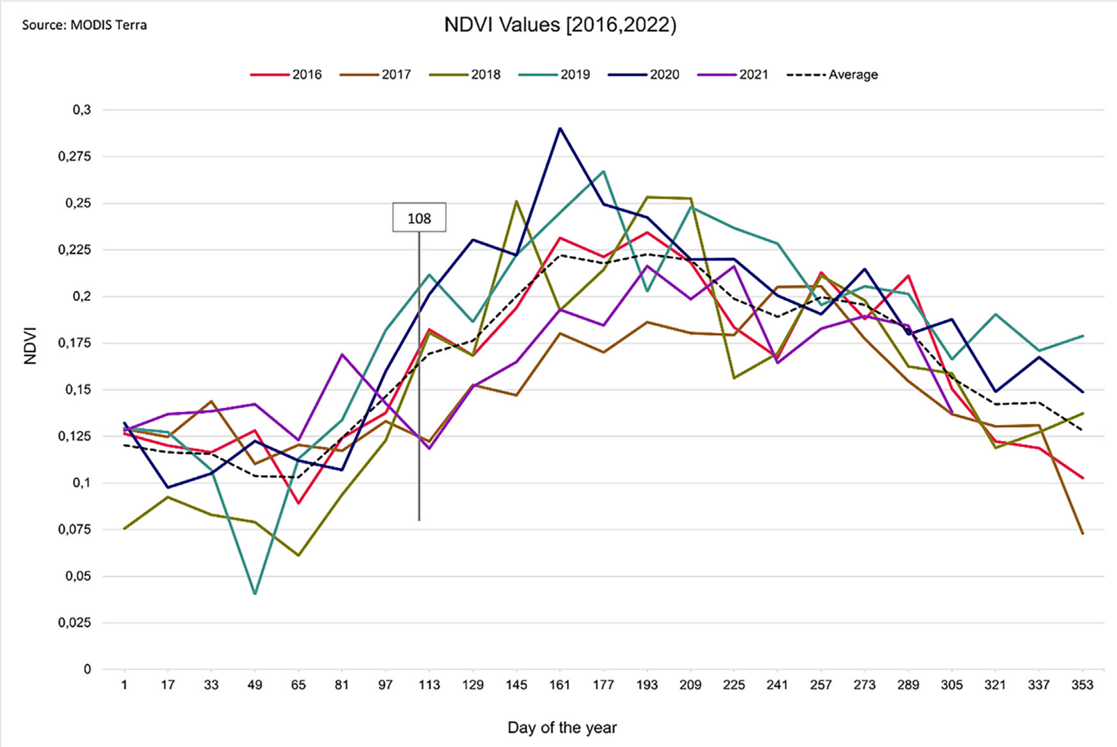

Fig. 1. MODIS NDVI by day of year over a six-year period (Source: Own compilation).

Results

Pre-wildfire conditions

Fig. 1 highlights the trends of Normalised Difference Vegetation Index (NDVI) Moderate resolution image spectra radiometer (MODIS) values over a six-year period (2016 -2021) for the Cape Peninsula region. The relevance of NDVI is that it strongly correlated to vegetation moisture and thus is a good proxy indicator for climate conditions as demonstrated by the 2017 trend line (dark brown line) which was the driest year in the well documented 2015-2018 drought of the Western Cape. More emphasis will be placed on the 2021 trend as it is the time leading to the fire case discussed in the paper. Fig. 1 which is grouped by day of the year allowed the researchers to reconstruct the conditions leading up to fire on April 18th, 2021 (year day 108) to be compared to the same period in previous years. From day 0 (the 1st of January) to day 81, 2021 NDVI values are prominently higher relative to all six previous years. NDVI values at day 0 at the start of the 2021 year were like previous years (except 2018 which was the end of a drought period) but then increases. From day 65 to day 81, an increase in values was noted, most likely indicating a short, wet period. After day 81 there was rapid and continuous decrease in values, well below average, until day 108, the day the wildfire began. Naturally after the wildfire, NDVI values in the region remained well below the average trend for the rest of the year as the vegetation was burnt.

Dam water levels were also found to be a significant indicator to consider due to the positive correlation between water storage and fire occurrence, that can provide cues on optimum fire conditions in an area. When analysing dam storage levels, it can be noted firstly that year-on-year levels have steadily increased since 2017 (the worst year of drought) and secondly that 2021 levels are noticeably higher throughout the year when compared to previous years and that surface levels are comfortably categorized as high with respect to the time of year. This holds true for the first in first 108 days of 2021.Observing the behaviour of maximum temperatures for a week leading up to the fire, the 11th to the 18th, show an increase when compared to the beginning and end of the month indicating a fleeting period of hotter than usual conditions. The top two hottest temperatures in the month, 30°C and 31°C, were recorded, right before ignition, on the 16th, 17^{th,} and 18th. Furthermore, MeteoBlue[8], highlighted that relative humidity for the month of April was largely consistent but a rapid and significant decrease is noted between the 16th and 18th of April. Similar conditions were reported by [4] in Northern Canada. Their findings show an increase in surface temperatures before the fire ignition date at burnt sites.

8 https://www.meteoblue.com/en/weather/week/cape-town_south-africa_3369157.

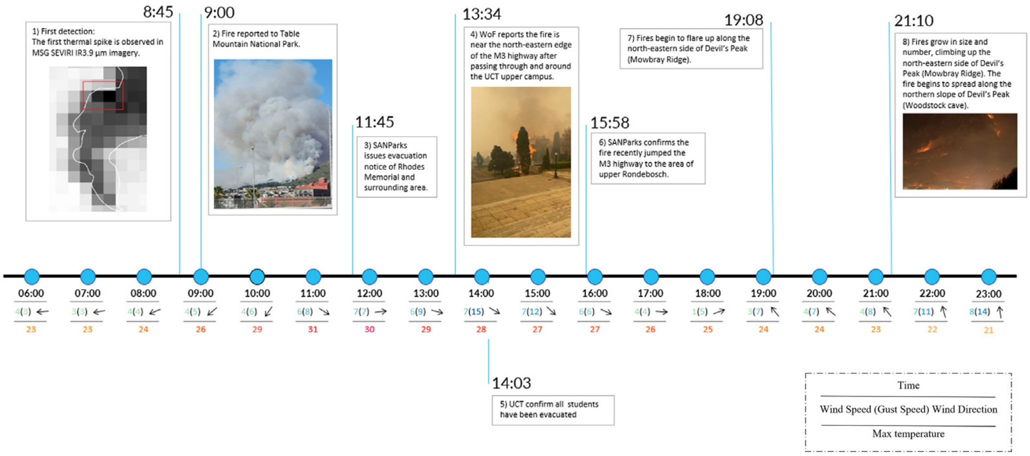

Fig. 2. First section of the wildfire timeline reconstruction using Twitter data (Source: Own compilation).

Fire timeline reconstruction from unstructured (Twitter) data

Based on the first thermal signature of the wildfire in MSG SEVIRI imagery taken at 8:45 ante-meridian (AM), it is estimated the fire began around 8:40 AM on the morning of the 18th as seen by feature one in the timeline represented, the wildfire was reported at 9:00 AM. As observed by the wind speeds and temperatures, the morning was characterized by hot conditions with a gentle northeasterly breeze. As the fire grew, the gentle breeze blew the fire towards Rhodes Memorial. At 10:00 AM the wind rapidly changes direction blowing towards the southeast and temperature increases to 29°C. 11:00 AM sees the hottest temperature of the day at 31°C and wind speeds are still relatively low at 6 knots with gusts of 8 knots. At 11:45 SANParks issue an evacuation of the Rhodes Memorial area indicating the imminent arrival of the wildfire in the area. By 13:34 the westerly wind has blown the fire down the slope from Rhodes Memorial into the forested region of the University of Cape Town (UCT) upper campus which acts as the WUI. From 12:00 PM to 15:00 past meridian (PM) sustained wind speed stays gentle while gusting wind speeds show significant increases from 7 to 9 to 12 to 15 knots. By 15:58 PM SANParks confirms the wildfire has spread over the M3 highway into upper Rondebosch. From 16:00 to 19:00, wind speeds and gusts are moderate. At 19:00 however the wind direction changes again from a northwesterly to a southeasterly wind. The shift in wind direction then contributed to fires beginning to spread up the northeastern side of Devil's Peak around 19:00 PM. By 21:00 fires moved over the northeastern edge onto the northern face of Devil's Peak. By 23:00 wind speeds had picked up to 8 and 14 knots sustained and gusting respectively.

Wind speeds continued to increase in the early hours of the 19th of April as the fire moves in a northwesterly direction both above and below Tafelberg Road towards Vredehoek at approximately 13:00. By 05:00 ante-meridian (AM) wind speeds are moderately strong but with significant gusts at 25 knots. The stronger wind drove the fire towards Vredehoek and by 07:00 AM gusts have peaked at 30 knots. At 09:22 AM Working on Fire (WoF) indicated that the fire was near Vredehoek. From mid-morning to mid-afternoon wind speeds gradually decreased but the fire burns up and along the northwestern slope of Devil's Peak. At 14:00 the fire was near Deer Park, and it was also noted that fire flare ups occurred in the unburnt regions of the northern slope of Devil's Peak causing fires to spread in the direction of the origin ignition site. Wind speeds continued to decrease into the early evening to 8 knots as the fire traveled up the northwestern slope both above and below Tafelberg Road (above Deer Park). At 22:42 it was observed that fire flares up above Deer Park. By midnight, fire still burns above and below Tafelberg Road. In the early morning, wind speeds indicated a gentle breeze, and the fire intensity appears to drop off. By 09:52 it was reported that the fire is mostly contained and remaining firefighting efforts were centered on Deer Park. By 13:30 it was reported that the fire had been contained. Figs. 2 and 3 below are highlights of the timeline and three-dimensional (3D) reconstruction of the fires progression and extent respectively as created by both the tweets and the described supporting data (image, weather and supporting auxiliary data).

Post-fire damage assessments

Burn scar mapping and Burn severity

Random forest classification yielded a burnt area of 5.31 km^2 which makes up 31.46% of the study region. The associated user's and producer's accuracy was 96.77% for both the burnt area and unburnt area class and the classification had an overall accuracy of 96.42%. The high overall accuracy achieved in this study compares well with reports from previous studies. The burnt and unburnt classes' accuracy metrics are comparable to [22] who achieved 98% and 97% in their binary classification of burnt areas in the southern Western Cape. With respect to the variable importance metrics calculated, our

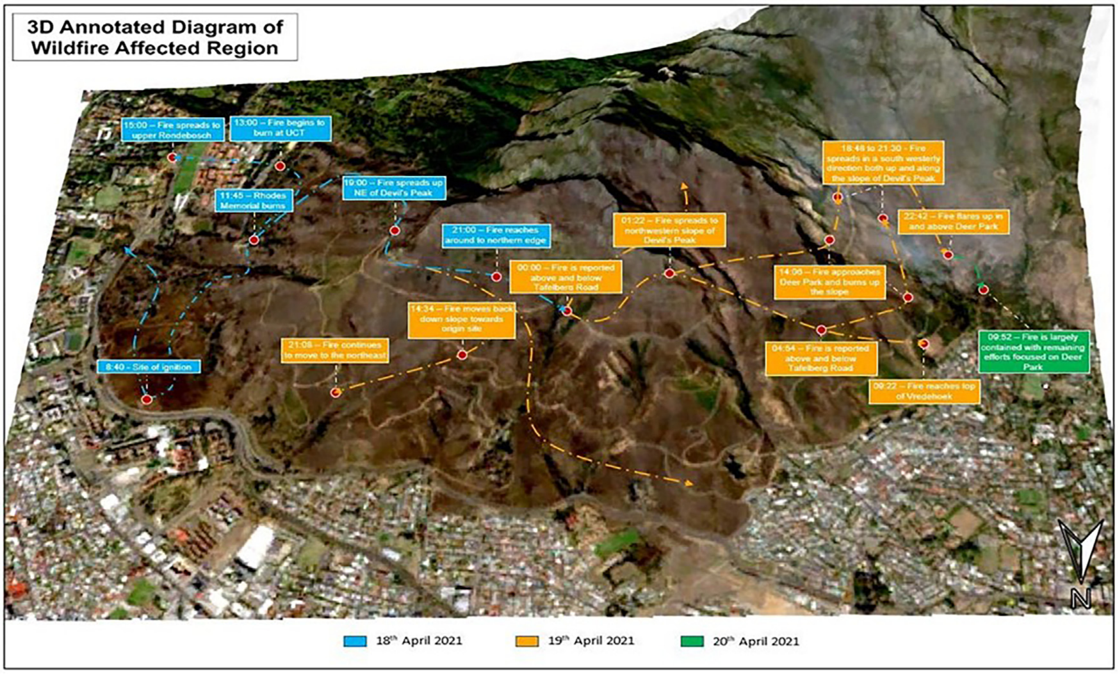

Fig. 3. First part of the wildfire timeline reconstruction using Twitter data (Source: Own compilation).

study found that the top five most influential variables in classification were NBR, B8 (NIR), B11 (SWIR 1), B6 (Red Edge 2) and NDVI.

Of the burnt area, 2.499 sqaure kilometers (km^2) (34.52%) was burnt, making it the largest burn severity class. Moderate-high severity was the next dominant class with 2.155 km^2 (30.37%). Moderate-low and low were the smallest classes with 1.316 km^2 (18.56%) and 1.174 km^2 (16.55%) respectively. The dNBR results show that most burnt regions are categorised as high and moderate to high severity.

When overlaying the burnt area with LULC data it was noted that the most affected land cover type was Shrubland with 3.32 kilometers (km^2) (46.84%) of burnt area. This class includes karoo and fynbos shrubland which is the dominant vegetation in the region. The large area burnt shows agreement with (Mucina et al. 2006) that fynbos is an ideal fuel source for fire. Forested Land made up 3.07 Square kilometers (km^2) (43.29%) of burnt area and built-up environments covered 0.48 km^2 (6.79%). Of the Shrubland burnt area, the highest severity class was moderate-high making up 42.94% whilst the highest severity class in the burnt Forested Land was 54.71%. This perhaps indicates that the fuels in Forested Land have higher combustibility than those in Shrublands.

Using a highly accurate random forest classification generated in Google Earth Engine, it was determined that 5.31 km^2 of land was burnt by the fire. As already highlighted, The dNBR index with land cover data determined that the land cover classes most affected by the fire were Shrubland (which includes the fynbos vegetation type) and Forested Land (making up 46.84% and 43.29% respectively of the burnt area identified by dNBR). Furthermore, of the burnt region 34.52% and 30.37% of it was classified as highly and moderately high respectively, indicating the severity of the April 2021 fire. The results from the random forest supervised classification burnt-unburnt map created an impression of the affected region as seen in Fig. 4 which illustrates a 2D classification map to assist in visualizing the affected area.

UAV based scene reconstruction of damaged building infrastructure

3D-SfM models were derived by the researchers to speak more to infrastructural damages. Due to the nature of the UAV survey, only exterior damage to buildings could be observed by the researchers and measured using the tools in Open Drone Map (ODM). No comment could be made by the researchers in this study about interior damage. Most of the affected infrastructure was contained within the University of Cape Town Upper Campus and is summarised as follows.

The University of Cape Town Jaggery library was the most impacted by the fire with 641,17 sqaure meters (m^2) of roofing affected and more than 89% of it being completely burnt through, exposing the interior of the building. The H.W Pearson building was impacted with 410,32m^2 of roofing and exterior burnt which totally exposed the interior of the top floor of the building which was also completely burnt. The University of Cape Town Upper Campus Residence and Fuller Hall were the next most affected buildings with relatively small area of roofing damage observed, 45,28m^2 and 27,71m^2 respectively. The damage was primarily partially burnt roof tiles and did not penetrate through the roofing (See Fig. 5). The damage to the Upper Campus Residence was exclusively on the roof of the western building while Fuller Hall had damage to the roof of the

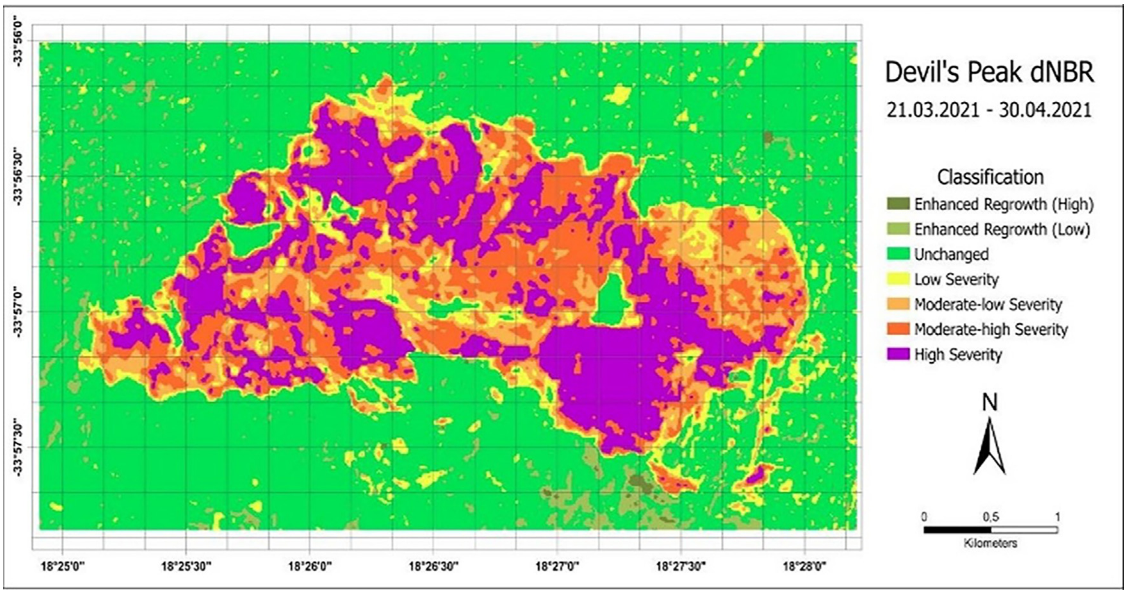

Fig. 4. A compilation of two-dimensional (2D) burnt area classification and burn severity classification (Source: Own compilation).

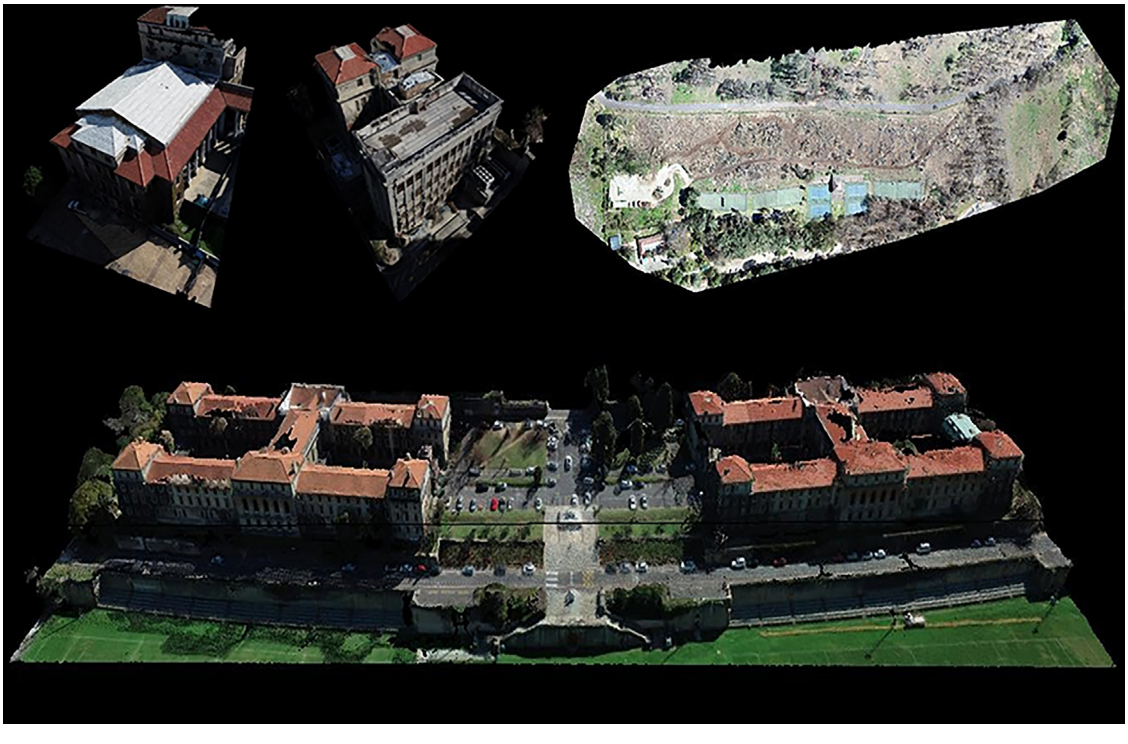

Fig. 5. a, b, c, d) 3D-SfM reconstructions of the Jagger Library, H.W Pearson building, Fuller Hall, and Upper Campus residence, respectfully. e) Constructed orthophoto of the burnt and cleared forests of the Upper Campus.

western building and damage to the decorative, memorial structure above the main entranceway into the building. Lastly, as noted in the wildfire timeline reconstruction, the trees on the western boundary of the study area, near the historical Rhodes Memorial were greatly impacted. As part of recovery effort and for future fire management many of these trees, both burnt and unburnt, were cut down. From the model of the area, it was measured that 3.965 km^2 of forestry had been affected and removed, effectively shifting the wildland-urban interface (WUI) away from the Upper Campus which is ideal for future wildfire management.

Discussion and conclusion

The study contributes towards meeting the Africa's Union's Agenda 2063 and the United Nations Sustainable Development Goals (2030) in curation, documentation and mitigation of community and environments. Based on multiple data sources, the study showed that unstructured datasets have a place in mainstream science. This section shares key discussions, conclusions, and recommendations of the study from pre-fire wildfire conditions, fire timeline reconstruction from unstructured (Twitter) data to post-fire damage assessments respectively (including burn mapping, burn severity and UAV based scene reconstruction of damaged building infrastructure).

The remotely sensed NDVI values derived from MODIS data and surface storage levels depicted the longer-term climate trends while weather data highlighted the short-term prevailing pre-fire conditions. The results showed that in months prior to the fire, climate conditions (and thus vegetation condition) were wetter than in the previous five years and objectively classified as good. However, in the two weeks leading up to the fire, vegetation health rapidly drops off and hot dry conditions prevailed. This is emblematic of a berg wind changing wind directions from a north-westerly to a westerly to a south easterly direction which could have provided significant influence on the fire breakout or spread from the point of origin to upper Rondebosch and then up and around Devil's Peak towards Vredehoek and Deer Park. It is thus noted that despite longer term trends, short term weather can have profound effect on fuels and ignition.

Via mapping unstructured Twitter data in conjunction with weather forecasts, a timeline of the fire spread was derived. The timeline combined structured and unstructured data to provide a combined visual that made it easier to explain the phenomenon causing the spread of fires. It can be noted that the high number of tweets on the second day, explains the big size of the fires and the large areal coverage of day 2. This can be justified to reflect a greater public response on day two that was prompted by the size of the fire on this day hence a greater need to tweet about it. This also resonated with a study conducted in North America, that reported a high number of emergency phone calls during the periods when the fire was most turbulent and close to certain locations[9]. As highlighted above, establishing the timeline of the fire from Twitter data with associated wind conditions, allowed for the spread path to be tracked post fire. Most importantly, tweet location information allowed the documentation of the influence of weather conditions and topography on the spread of fire to be added to the timeline. Additionally, the usage of Twitter data proved to be effective, particularly in curating an event in an urban area of South Africa as it allowed for additional 'on the ground' data with high temporal resolution. Social media is a rich source of information that has proved to be valuable in the tracking of fire events as it is often the first public communication channel for emergency services. However, due to the sheer volume of information on social spaces, it can also be difficult to find specific information on an event of interest resulting in a labour-intensive process and search algorithms can optimise that aspect.

The burnt severity map overestimated the area of the burnt region in comparison to the highly accurate burnt area classification. This can be attributed to the simplistic nature of the dNBR index as many unburnt urban areas and barren land were incorrectly labelled as low burn severity which was also the least accurate class. This study showed an over estimation of the burnt area by the dNBR index when compared to the burnt area classification particularly with the low severity class. This over estimation can be attributed to the simplicity of the index and that urban environments have a variety of surfaces and are heterogenous and prone to misclassification by spectral indices.

The very high-resolution imagery acquired by a UAV and 3D modelling facilitated graphical assessment of affected property and infrastructural damage measurements. The modelling results discussed in section 4 and depicted in Fig. 5 highlighted the most affected buildings and accounted for the April 2021 fire event damages. The 3D models and orthophotos constructed from a SfM photogrammetry process of the affected regions and buildings allowed for exterior damage to be measured with less detail on interior damage as that would have required indoor mapping to curate internal damage.

The fusion of cloud-based platforms like Google Earth Engine (GEE) with other structured and unstructured spatial datasets for fire analysis, particularly in terms of in classification was appreciated in this study. The study illustrated the importance of remote sensing and weather station data in documenting wildfire events. It also depicts the overall significance of fusing unstructured and structured for geospatial applications and the role it can play in closing the voids in wildfire reporting. It is recommended that future studies test the burn severity classification using machine learning derived classification methods like Random Forest classification and continue to take advantage of rich tools and datasets such as GEE to curate historical fires in South Africa. Finally, in future investigations of wildfires, it may be valuable to leverage the full power of the paid Twitter Application Programming Interface (API) service which gives far greater control concerning filtering out unwanted data and provides valuable statistical analytics of trends and other metrics.

Accurate documentation and mitigation of wildfires is therefore significant in addressing Sustainable Goal eleven (11) as well as the African Unions Agenda 2063 Goal one (1) which resonate an intention to preserve sustainable environments, communities, and cities for future generations. Through documentation of wildfires, long-term fire management, legal investigation and historic archives can be curated. There is genuine motivation to further explore the novel integration of unstructured social media into fusing unstructured data with structured data for geospatial data processing.

[9] https://www.nfpa.org/News-and-Research/Data-research-and-tools/Emergency-Responders/Fire-department-calls.

Ethical Statement

No human or animal subjects participated in the study.

Author contributions

Daniel O'Sullivan Hewlett conceptualised the study, collected data, conducted investigations, drafted the methodology to produce results and analysed the data presented in the project. Moreblessings Shoko provided leadership, conceptualisation, supervision, project administration and review writing for public dissemination, while Brighton Chamunorwa assisted with writing with a focus for publication.

Declaration of Competing Interest

The authors solemnly declare no conflict of interest.

Acknowledgements and Funding

The Researcher acknowledges EBE Talks, EBE Research fund and Dr Walter Uys for funding and support.

References

[1] R.S. Allison, J.M. Johnston, G. Craig, S. Jennings, Airborne optical and thermal remote sensing for wildfire detection and monitoring, Sensors 16 (8) (2016), doi:10.3390/s16081310.
[2] V.G. Ambrosia, S.W. Buechel, J.A. Brass, J.R. Peterson, R.H. Davies, R.J. Kane, S. Spain, An integration of remote sensing, GIs, and information distribution for wildfire detection and management, Photogramm. Eng. Remote Sens. 64 (10) (October 1998) 977–985 https://www.asprs.org/wp-content/uploads/pers/1998journal/oct/1998_oct_977-985.pdf.
[3] A. Bouguettaya, H Zarzour, A.M. Taberkit, A. Kechida, A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms, Signal Process. 190 (2022), doi:10.1016/j.sigpro.2021.108309.
[4] M. Carmo, F. Moreira, P. Casimiro, P. Vaz, Land use and topography influences on wildfire occurrence in northern Portugal, Landsc. Urban Plan. 100 (1–2) (2011) 169–176, doi:10.1016/J.LANDURBPLAN.2010.11.017.
[5] W. Chen, Y. Zhou, E. Zhou, Z. Xiang, W. Zhou, J. Lu, Wildfire risk assessment of transmission-line corridors based on naïve Bayes network and remote sensing data, Sensors 21 (2021) 634, doi:10.3390/s21020634.
[6] L. Collins, P. Griffioenb, G. Newell, A. Mellor, The utility of random forests for wildfire severity mapping, Remote Sens. Environ. 216 (2018) 374–384 https://ui.adsabs.harvard.edu/link_gateway/2018RSEnv.216..374C/, doi:10.1016/j.rse.2018.07.005.
[7] A.M. Cunliffe, R.E. Brazier, K. Anderson, Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from motion photogrammetry, Remote Sens. Environ. 183 (2016) 129–143, doi:10.1016/J.RSE.2016.05.019.
[8] F.A. Engelbrecht, J.L. McGregor, C.J. Engelbrecht, Dynamics of the conformal-CubicFun atmospheric model projected climate-change signal over southern Africa, Int. J. Climatol. 29 (7) (2009) 1013–1033, doi:10.1002/joc.1742.
[9] Frost, P. E., Kleyn, L. G., van den Dool, R., Burgess, M., Vhengani, L., Steenkamp, K., & Wessels, K. (2018). The Elandskraal Fire, Knysna., https://www.researchgate.net/publication/326112117_The_Elandskraal_Fire_Knysna_A_data_driven_analysis
[10] D. Fornacca, G. Ren, W. Xiao, Evaluating the best spectral indices for the detection of burn scars at several post-fire dates in a mountainous region of Northwest Yunnan, China, Remote Sensing 10 (8) (2018) 1196, doi:10.3390/RS10081196.
[11] R.A. Galantucci, F. Fatiguso, Advanced damage detection techniques in historical buildings using digital photogrammetry and 3D surface analysis, J. Cult. Heritage 36 (2019) 51–62, doi:10.1016/J.CULHER.2018.09.014.
[12] G. Jan Van Oldenborgh, F. Krikken, S. Lewis, N.J. Leach, F. Lehner, K.R. Saunders, M. Van Weele, K. Haustein, S. Li, D. Wallom, S. Sparrow, J. Arrighi, R.K. Singh, M.K. Van Aalst, S.Y. Philip, R. Vautard, F.E.L Otto, Attribution of the Australian bushfire risk to anthropogenic climate change, Nat. Hazards Earth Syst. Sci. 21 (3) (2021) 941–960, doi:10.5194/NHESS-21-941-2021.
[13] M. Kganyago, L. Shikwambana, Assessment of the characteristics of recent major wildfires in the USA, Australia, and Brazil in 2018–2019 Using Multi-Source Satellite Products, Remote Sens. 12 (11) (2020) 1803, doi:10.3390/RS12111803.
[14] B. Leblon, L. Bourgeau-Chavez, J. San-Miguel-Ayanz, Use of remote sensing in wildfire management, in: Sustainable Development-Authoritative and Leading-Edge Content for Environmental Management, 2012, pp. 55–82, doi:10.5772/45829.
[15] S.A. Morshed, K.M. Ahmed, K. Amine, K.A. Moinuddin, Trend analysis of large-scale twitter data based on witnesses during a hazardous event: a case study on california wildfire evacuation, World J. Eng. Technol. 9 (2021) 229–239, doi:10.4236/wjet.2021.92016.
[16] Mucina L., Rutherford M.C., Powrie L.W., Ellis F., Fynbos Biome, (2006), https://www.researchgate.net/publication/40106510
[17] M. Syifa, M. Panahi, C. Lee, Mapping of post-wildfire burned area using a hybrid algorithm and satellite data: the case of the camp fire wildfire in California, USA, Remote Sens. 12 (2020) 623, doi:10.3390/rs12040623.
[18] G.R. Van Der Werf, J.T. Randerson, L. Giglio, G.J. Collatz, M. Mu, P.S. Kasibhatla, D.C. Morton, R.S. Defries, Y. Jin, T.T. Van Leeuwen, Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997-2009), Atmos. Chem. Phys. 10 (23) (2010) 11707–11735, doi:10.5194/ACP-10-11707-2010.
[19] S. Vieweg, A.L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events: what twitter may contribute to situational awareness, CHI 2010: Crisis Informatics, April 10–15, 2010 Atlanta, GA, USA ., doi:10.1145/1753326.1753486.
[20] Z. Wang, X. Ye, Social media analytics for natural disaster management, Int. J. Geograph. Inf. Sci. 32 (1) (2018) 49–72 VOL.NO., doi:10.1080/13658816.2017.1367003.
[21] Y. Xie, Z. Sha, M. Yu, Remote sensing imagery in vegetation mapping: a review, J. Plant Ecol. 1 (1) (2008) 9–23, doi:10.1093/jpe/rtm005.
[22] S. Xulu, N. Mbatha, K. Peerbhay, Burned area mapping over the southern cape forestry region, south africa using sentinel data within GEE cloud platform, ISPRS Int. J. Geo-Information 10 (8) (2021) 511 2021, Vol. 10, Page 511, doi:10.3390/IJGI10080511.
[23] X. Zhou, P. Wang, K. Tansey, S. Zhang, H. Li, L. Wang, Developing a fused vegetation temperature condition index for drought monitoring at field scales using Sentinel-2 and MODIS imagery, Comput. Electron. Agric. 168 (2020) 105144, doi:10.1016/J.COMPAG.2019.105144.