Proceedings World Geothermal Congress 2023
Beijing, China, April 17 – 21, 2023

.

Detection and monitoring of thermal anomalies using Satellite-based thermal infrared time series data in the Eburru Geothermal prospect area, Kenya

Bernard Omwenga., Eunice Bonyo

Bomwenga@kengen.co.ke

EBonyo@kengen.co.ke

Keywords: Time series, thermal anomalies, monitoring

ABSTRACT

Different material on the surface of the earth have a characteristic thermal anomalies. By measuring these thermal anomalies, we can locate areas with high temperature contrasts. Geothermal environments are usually characterized by high temperature anomalies due to surface manifestation features such as fumaroles, hot grounds, and steam vents. As such, information derived from such measurements helps in the comprehension of mass flow, degassing regimes, and changes in barometric pressures associated with geothermal systems. This makes it possible to characterize geothermal systems. Different methods are used for thermal measurements of thermal features. Field-based techniques are the most common techniques used for such studies. However, this approach is limited in scope, with regard to repeatability, and in some instances inability to access zones with rugged terrains. The availability of long-term global satellite data (imageries) offers an alternative way of thermal measurements. The method is convenient due to its ability to estimate land surface temperature over large areas of coverage, in a relatively short time. By monitoring the fluctuation of temperature values of specific pixels, over time it is possible to understand the thermal history of specific hotspot zones, within an area of interest. For this study, thermal bands of the multispectral Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) images acquired between 2018 and 2022, were tested at known geothermal hotspot zones within the Eburru Geothermal prospect area. In addition, specific points from non-geothermal zones (e.g., known vegetated areas and zones with normal surface temperatures) were used as control points. Emphasis was placed on night time scenes, to reduce the effects of solar illumination. The amplitudes of the background removed time series plots show stationarity over time, an indication that minimal changes have occurred in geothermal hotspot zones. However, the time series curves of pixels from non-geothermal areas exhibit inconsistent fluctuation of the peaks and outlier signals, an indication that other factors (mainly weather) affect their thermal behavior. Future studies should incorporate time-series data from ground measurements from heat loss survey measurements to for further validation. In addition, hypertemporal thermal data (preferably acquired by stationary thermal infrared cameras) will be useful in the precise monitoring of the surface thermal behavior in the area.

1 INTRODUCTION

Geothermal surface manifestations features include fumaroles, hot grounds, and mud pools. In many cases, these features have notably elevated temperatures compared with their surroundings. The elevated temperatures are identified as thermal anomalies which result from the emission of electromagnetic radiation from their surfaces (Saraf et al., 2009). The detection of these anomalies is useful for (1)-quantification and prospecting of geothermal resources (Haselwimmer et al., 2013), (2)-identification of hazard precursors e.g., volcanic eruptions (Chiodini et al., 2016), and (3)-earthquakes prediction based on an assessment of elevated land surface temperature due to subsurface stress build-up (Pavlidou et al., 2016; Saraf et al., 2009).

In geothermal exploration, the focus has been on the detection of thermal anomalies for resource identification/prospecting. Conventionally, field-based thermal detection techniques (chloride inventory and direct soil temperature measurements) use thermometers to determine the heat flux in fissured geothermal areas. These measurements are instrumental for understanding and estimating the energy potential and capacities of geothermal reservoirs (Bloomberg et al., 2014, Frondini et al., 2004). Recently, attention has been shifting to monitoring and understanding the long-term sustainability of geothermal resources in a given area. As such, remote sensed thermal infrared data (TIR) has emerged as one of the valuable tools for monitoring soil temperature gradients in active geothermal areas (Mia et al., 2012).

Spaceborne TIR products are preferred because most of their data is open–sourced, have a large swath guaranteeing a large area of coverage, and have higher temporal frequency, hence applicable for long-term temporal analysis. The common metrics for quantification of TIR products, such as land surface temperature (LST), and sea temperature (Dech et al., 1998), have become an important tool in assessing the surface temperature dynamics. These product metrics have been used in the monitoring of thermally anomalous zones in active geothermal fields (Mia et al., 2012). However, these measurements are significantly affected by external influences associated with seasonal and diurnal weather patterns (Vaughan et al ., 2012).

Therefore, this study aims to develop a methodology to eliminate external influences in TIR and analyze the resultant temporal fluctuations to detect outlier signals when assessing changes in a geothermal field. We investigate the presence of extreme outlier signals in the time series data acquired by the Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) to explore the temporal pixel temperature fluctuations. We assess the correlation of the outliers with weather and their potential as tools for detecting changes in the hydrothermal system. We apply the method to dictate thermal anomalies in Eburru geothermal field in Kenya from ECOSTRESS data acquired between August 2018 and July 2022.

1.1 Specific objectives

1. To identify and isolate imageries with low pixel temperature values resulting from atmospheric conditions (clouds).
2. To analyze the spatio-temporal development of the pixel temperatures isolated from the background of the study area.
3. To identify, and analyze outlier signals of TIR time series data (corrected for cloud cover), and compare them with the anomalous values of ambient localized weather data; with explicit emphasis on where, how, and when they occur.

3.1. Hypothesis

H_1: Short-term abrupt anomalies are related to localized diurnal weather effects, while long-term incremental changes characterized by non-stationary time series of fumarole temperature is associated with episodes of hydrothermal unrest, or decline of heat output (heat flux) of the hydrothermal manifestation features (Chiodini et al.,2011).
H_2: High correlation (coefficient of determination (r^2) >=0.6) between pixel temperature and meteorological data (surface moisture, ambient temperature) indicates that localized meteorological factors affect the temperature variation of thermal infrared signals from year to year (Malvina et al., 2019).

1.2 Study Area

The Eburru geothermal field is an active geothermal system located in the central segment of the Kenya Rift valley bordering Lake Naivasha to the South and Lake Elementaita to the North (Figure 1). The geothermal resources exploitation in Eburru is used for power generation and other direct uses. The field has an estimated geothermal potential of 25 Mwe (Mwarania, 2014). Its proven exploited power capacity is 2.52 Mwe (Omenda, 2012).

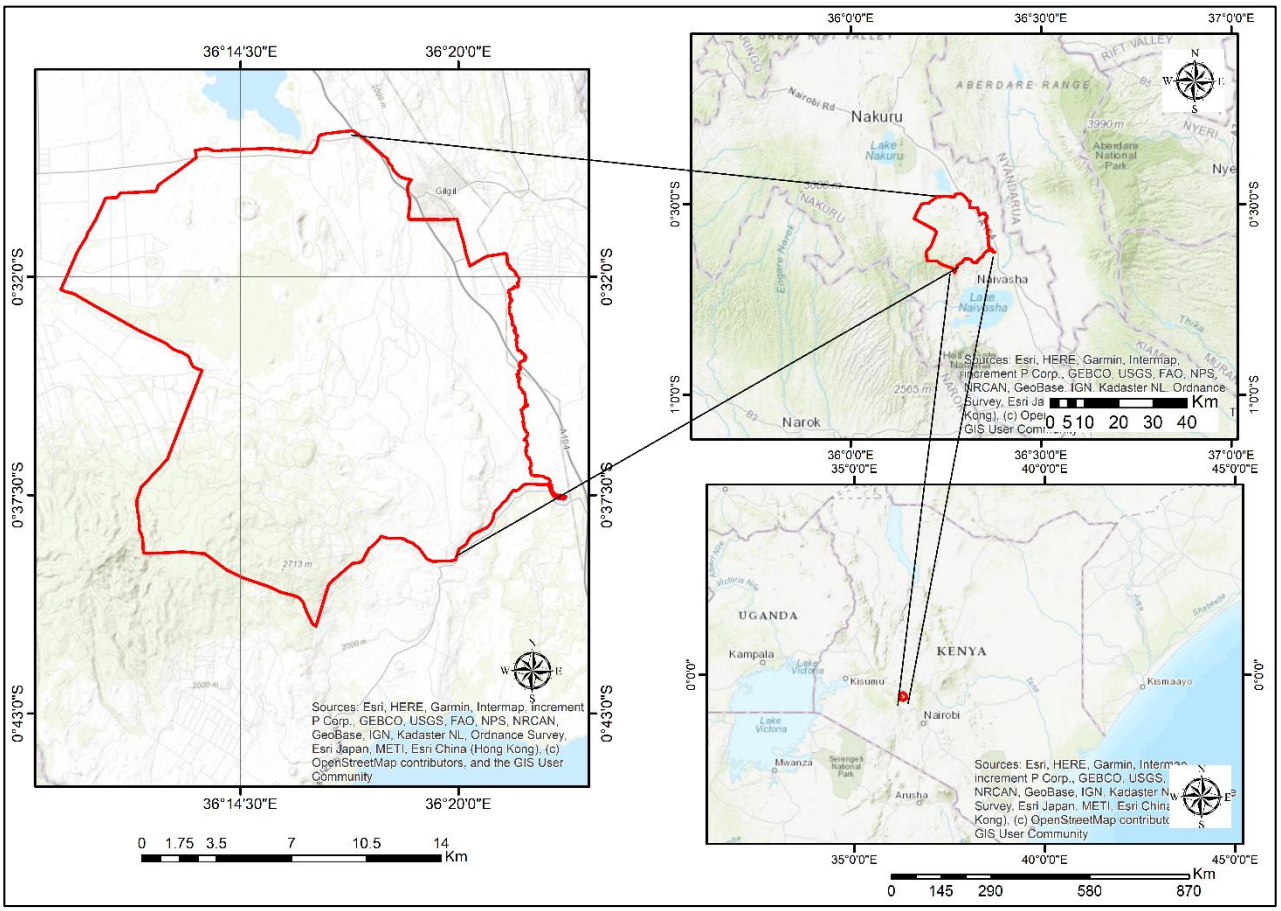

Figure 1: A map showing the geographical location of the area of study (Modified from Esri's world online topographic map)

The geology of Ebuuru is described as a caldera and fault-dominated system comprising Pleistocene to Holocene lithological units. The main rock units include pyroclastic deposits, pantelleritic trachytes, pantellerites commendites, and basalts (Clarke et al., 1990). The area has numerous surface geothermal manifestation features, including fumaroles, hot grounds and altered surfaces. These features are mainly aligned along major and minor geological structures (i.e., faults and fissures) and also exhibit anomalously high-temperature values.

2 DATA

2.1 TIR data

The main datasets used in this study are the images acquired by the Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS). The ECOSTRESS products provide Land Surface Temperature (LST) and surface emissivity data in five channels in the Thermal Infrared portion between 8 and 12.5 µm (Shi and Hu, 2021) at a 70m spatial resolution. The instrument has a large swath width of about 402 km and a temporal repeat of approximately 3-5 days, allowing intensive temporal thermal and spatial monitoring. These images include the level-2 (ECO2LSTE) products. It also has extra bands for geolocation (at 1.6 µm), and detection of clouds, existing in the ECO2CLD format.

Images with frames with a geographical extent of 60x60 km were downloaded from the NASA's Land Processed Distributed Active Archive Center website at https://e4ftl01.cr.usgs.gov/ECOSTRESS/. In this study, nighttime images were selected to minimize the solar effects on the recorded thermal signal (Kuenzer & Dech, 2013). In addition, the downloaded images were restricted to a minimum and maximum cloud cover between 5-20%. As a result, only 144 images that met the aforementioned criterion were selected between 2018 and 2022.

2.2 Meteorological data

Meteorological data used in this study include daily air temperature, soil moisture and rainfall. The meteorological data was downloaded from an open-source POWER Data Access Viewer website (https://power.larc.nasa.gov/data-access-viewer/). The website offers archived weather datasets of global dimension from 1981 todate. This dataset is acquired by the Meteosat Second Generation's (MSG) Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument. The instrument has a 3 km spatial resolution and a 15 minute revisit time.

The data is statistically correlated with ECOSTRESS pixel temperature data to ascertain their relation with exogenous changes. This offers a validation measurement as a proxy of the surface-atmospheric interaction (Malvina et al., 2019), known to affect the recorded radiant temperatures acquired by the spaceborne instruments.

3 METHODS

The method consists of four steps: pre-processing, extraction of temporal plots, analysis of time series plots, and anomaly detection.

3.1 Pre-processing of TIR data

The image pre-processing in this study entailed:

i. The correction of atmospheric absorption effects, notably the intense cloud cover. According to Coolbaugh et al., (2007), such artefacts dominate nighttime scenes and underestimate recorded radiant temperature values. The cloud masking algorithm by Baroncini et al. (2008) implemented using a python code (edited from NASA.org) was used (Appendix 1).
i. To further filter out anomalous pixels not flagged out by the algorithm, we set a threshold to remove these outlier pixels by calculating quantiles. In this step, pixel values lower than 2% quantile, or higher than 98% quantiles were flagged as anomalous and thus replaced by NaN values.
ii. In addition, we quantified missing pixels of the cloud-masked time series images and then excluded scenes with over 50% Not a Number (NaN) values.
iii. The other pre-processing steps include the conversion of radiant temperatures of the corrected ECOSTRESS scenes to kinetic energy using the Temperature/Emissivity Separation (TES) algorithms by Gillespie et al. (1998). In this phase, the temperatures were extracted from the raw ECOSTRESS data, by converting the data values to degrees Celsius using a scaling relation of $T = 0.02 * x - 273.5$ (proposed by Henderson et al. (2019).
iv. The final phase involved the subsetting of the corrected pseudo images along specific regions of interest to an approximate size of 5x5 m^2 for further analysis.

3.2 Extraction of temporal plots/profiles from the selected TIR images

This step involves the extraction of time series plots from two regions of interest with contrasting spatial and temporal thermal behaviour. The first anomaly is linked to pixels with relatively elevated land surface temperatures (above a specific user-defined threshold). These zones were majorly identified along hydrothermal manifestation features such as hot grounds, fumaroles, and major geological structures (faults, and fractures). Such zones are postulated to be subjected to geothermal effects. The second anomaly (background areas), includes pixels around cooler objects (non-geothermal regions). The location of these zones was mainly identified along densely vegetated zones, in the South Eastern part of the study area. Such regions have relatively low land surface temperature values. These zones were visually identified as shown in Figure 2. The second step involved the creation of training polygons around selected anomalous zones. Average pixel temperature values around the selected training polygons were thereafter extracted, and converted into ASCII, and comma-separated values (CSV) formats, for further analysis. A python script (Appendix 2), was used to perform this task.

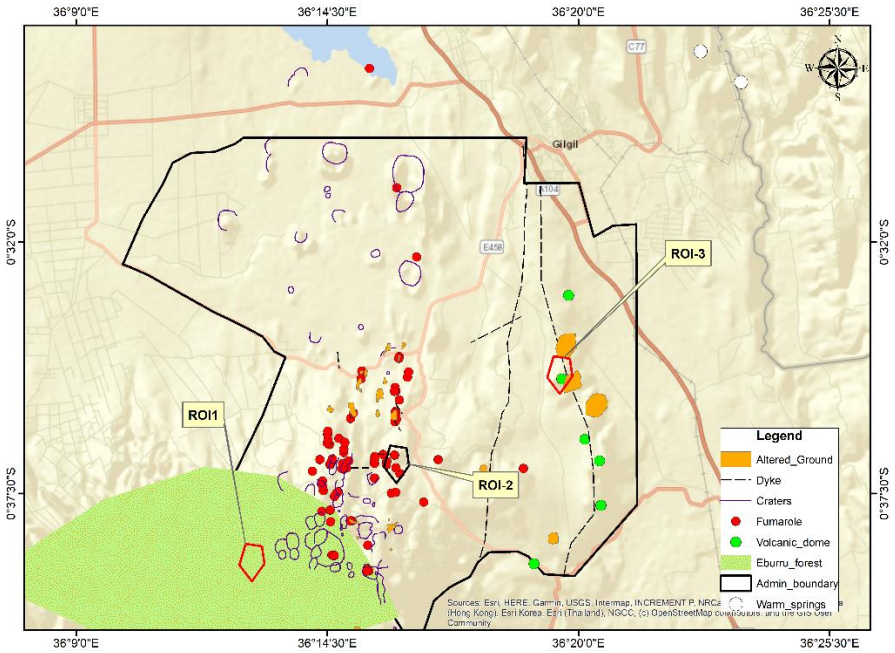

Figure 2: Map of the study area showing the location of hydrothermal alteration zones, and polygons characterizing three Regions of Interest (ROI-1, ROI-2, ROI-3), selected for thermal analysis. ROI-1, represents the forested area, ROI-2 symbolizes areas adjacent to fumaroles and altered grounds, and ROI-3 shows anomalous zones along major geological structures (i.e., fault lines, dykes, and ridges).

To further suppress temporal non-geothermal variation, the temperature above the background (ΔT), was calculated by subtracting the background temperature from a given pixel temperature value using the Murphy et.al. (2011) method as follows:

$$\Delta T_{i,k} = T_{i,k} - BT_k$$

Where $\Delta T_{i,k}$ is the temperature above the background, $T_{i,k}$ is the pixel temperature of the nth scene, BT_k is the background pixel of kth scene. The temporal fluctuation of the temperature above the background was monitored for the stationarity, and advent of anomalies as discussed in the subsequent subsections.

The temperature above the background (ΔT) time series plots were quantitatively and qualitatively analyzed for seasonality, and temporal variation. The quantitative analysis involved the visual inspection of the trends, frequencies, wavelengths, and amplitudes of the time series plots. A more robust Augmented Dickey-Fuller test (ADF) method (Atwan, 2022), was used to qualitatively test for the time series' data stationarity. In this method, if the p-value in ADF is less than the significant level (0.05), the hypothesis is rejected, hence the time series data is deemed stationary. The analysis was solely used in the preliminary or exploratory analysis phase.

3.3 Computation of thermal anomalies and trends

This step involves the detection of outlier data points out of a sample population with irregularly large deviations, which correspond to unexplained large variations beyond trend and seasonality. As hypothesized, such events correspond to signals associated with thermal or weather-related events. Four techniques were used to detect anomalies; (a) manual flagging of outlier points by visual inspection of the amplitudes of the time series plots (Saraf et al., 2009), (b) automated flagging of anomalous points using the deviation method implemented using a python code, available at https://github.com/ritvikmath/Time-Series-Analysis., (c) the Turkey fences method, and (d) z-score (or the standardized scores method) by Jimenez-Munoz et al.(2015).

In the deviation technique (Pastor et al., 2020), the standard deviation or means of the datasets is calculated on a per-month basis. The datasets of specific months were grouped and analyzed for unusual behaviour (abnormal amplitudes or kinks), which were thereafter flagged out. The procedure was repeated for the meteorological datasets (mean daily temperatures and soil moisture) of the specific months and matched to ascertain their quantitative similarities.

The Turkey Fences method (Atwan, 2022), involves the use of descriptive statistics to isolate values in the lower, and upper quantile in a distribution set. The most extreme values that, lie below the lower fence of the sample population are calculated as Q1 – (1.5 x IQR) or whereas those above the upper limit are calculated as Q3 + (1.5 x IQR). This method was used to validate the outlier points identified by the deviation and manual flagging techniques. The technique was implemented using a script developed in python platform.

The Standardized Normalized Variate (z-score) method shown in Equation xx, was used to calculate standardized scores, and establish the distance between each observation and the mean, to qualitatively rank the detected anomalous values observed in the time series plots.

$$ET_{std} = \frac{ET1 - ET2}{\sigma}$$

Where ET_{std} is the standardized score (z-score), ET1 is the original temperature (an observation), ET2 is the mean of the dataset, and σ (sigma) is the standard deviation of the dataset. The ranking criteria of standardized scores used (adopted from Jimenez-Munoz) to flag anomalous events are as follows: [+0.5 to +0.8]- Abnormal, [+0.8 to +1.3]- Moderate, [+1.3 to +1.6]-Severe, [+1.6 to +2]-Extreme, and [+ or higher)-exceptional. These scales were used to further categorize the pixel temperature above the background (Calculated in Equation xx). In this study, for pixel temperatures to be considered anomalously hot/cold, the values must exceed a threshold [0.6+2.0], or (>2).

3.4 Processing of meteorological data

Initial pre-processing involves resampling hourly data to a daily frequency to conform the ECOSTRESS pixel temperature data. Univariate imputation was performed in instances with null values (Atwan, 2022). In this technique, subsequent values before the NaNs are averaged, and the resulting values replace the missing data points. Once a coherent dataset was attained, a statistical correlation (Pearson correlation method) was performed, and the results were presented in a correlation matrix diagram. This was mainly performed using R scripts.

4 RESULTS

4.1 Atmospheric correction (Cloud masking)

The results of cloud masking of the raw ECOSTRESS images are shown in Figure 3. We significantly reduced background temperature noise in the final processed image (Figure 3-C) compared to the raw night temperature images (Figure 3 A). A significant amount of noise (majorly from the cloud cover) has been reduced by the cloud-masking algorithm, as shown by the NaNs pixels in the cloud masked file of Figure 3-C. This is confirmed by the visual and statistical assessment of the resultant image/ scenes,. In which the scenes' standard deviations reduced average between 53% and 34%.

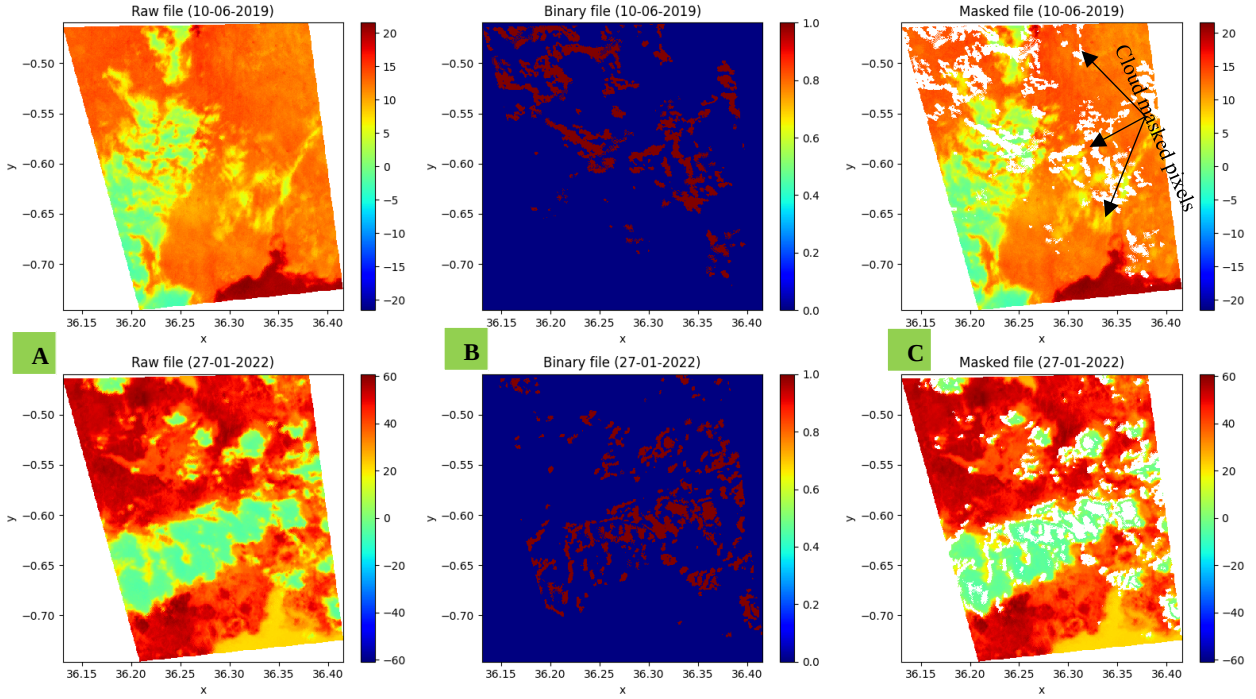

Figure 3: Sample ECOSTRESS images showing the distribution of surface temperatures, where (A):-original raw TIR scenes, (B): -a is a binary file showing areas with clouds (value 1), and without clouds (value 0). C: is the cloud-masked scenes acquired over the same period (2019-06-10 & 2022-01-27 at 04:37 UTC. The whitish segments in figure IC, show pixels with intense cloud cover (below 2% and above 98% quantile), masked by a user-defined algorithm.

4.2 Spatiotemporal distribution of surface temperatures

The purpose of this section is to examine and determine the spatial and temporal distribution of surface temperature (recorded by spaceborne platforms), and whether these variabilities are dependent on the land cover or material property of an area of interest. These findings were used to assess the overall trends, and whether they can insight into the sub-surface behavior in an active geothermal set-up. It is observed that the spatiotemporal extent of thermally anomalous zones in the Eburru Geothermal field is shown in Figure 4.

There is an uneven distribution of surface temperature in the focus area. The southern regions, areas around Eburru center (ROI1) which are characterized by numerous hydrothermal manifestation features, exhibit anomalous temperature values (hot anomalies). The anomalies are also evident in highly faulted and structurally infested regions in the eastern parts of the study area (ROI3), and the northern region along the Badlands basaltic flows. In contrast, the vegetated regions in the South Eastern part of the study area (ROI2), have low pixel temperature values but exhibit an incremental increase in surface temperature over the study period (positive temporal fluctuation).

The temporal plots of daily temperature values, extracted from pixels in specific Regions of Interest (ROIs) in Figure 4 (C), exhibit a seasonal pattern. They are also stationary (i.e., the properties such as mean/variance remain steady), as confirmed by the Dickey-Fuller statistical test results in Appendix 3, where the p values are less than 0.05. In addition, these time series plots are characterized by anomalous peaks (extreme negative and positive temperature values). The anomalous values are most pronounced, in 2022, with contrasting temperature values, elaborately expounded in section 4.3. Similar behavior is shown by the average temperature above the background plots (Figure 4 D), where significant outlier signals are discerned between 2018-2019. Such time series plots do not show significant changes in temperature values over time, owing to the stationarity of the time series plots.

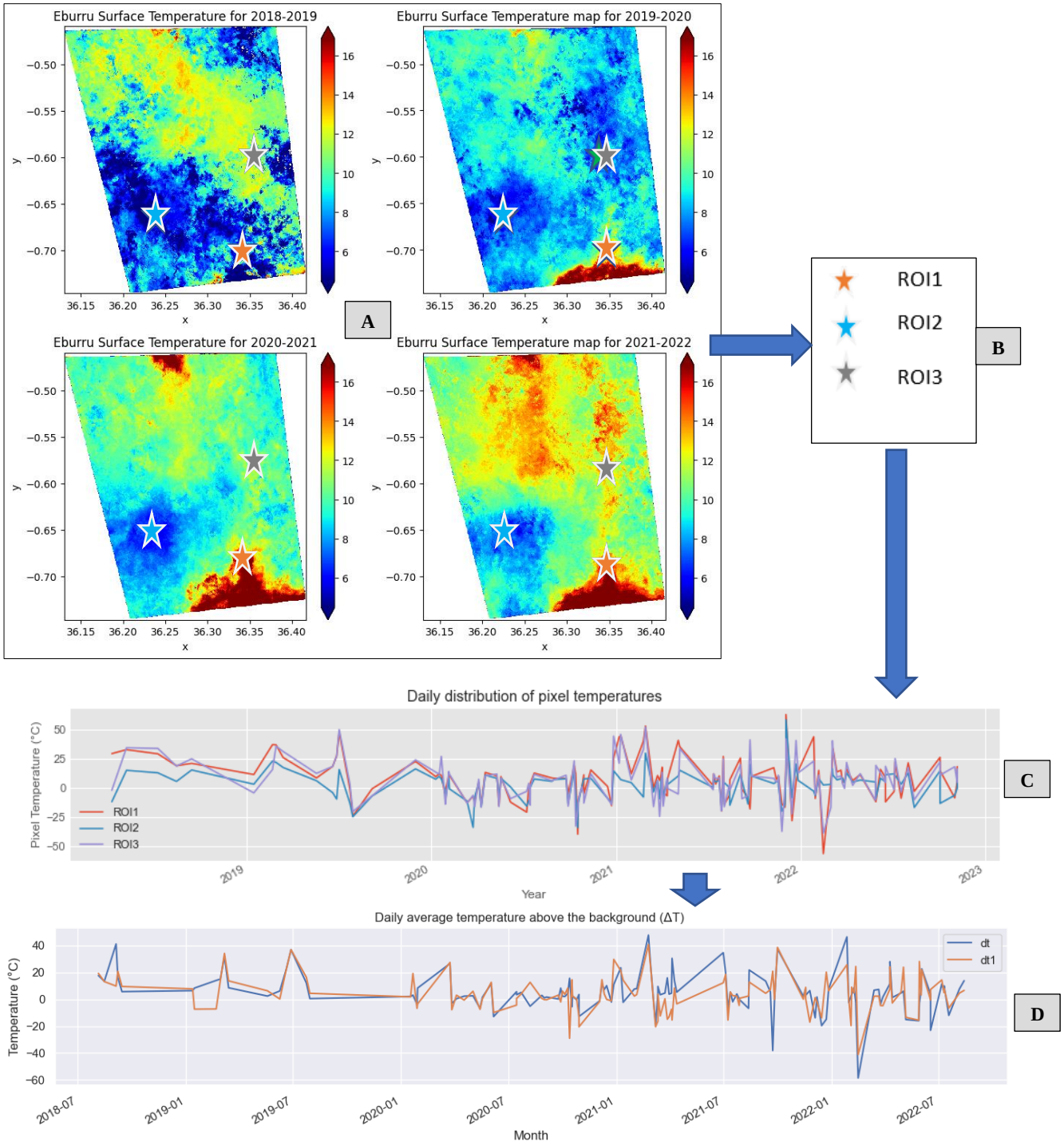

Figure 4:(A): Yearly instantaneous composite maps showing the spatio-temporal variation of the surface temperature of cloud-masked ECOSTRESS images from August 2018-July 2022, (B): -Regions of Interests along anomalous zones, C: Time series plot showing a variation of daily pixel temperatures and D: temperature above the background (ΔT) of respective ROIs, extracted from ECOSTRESS time series images.

4.3 Temporal extent (detection) of thermal anomalies and comparison with meteorological data

The purpose of this section is to determine the presence of anomalies and ascertain whether they are controlled by meteorological factors or hydrothermal unrest. The manually flagged anomaly plots (Figure 5 A & B) created by visual inspection of amplitudes of

the temporal plots of temperature above background (ΔT) reveal notable anomalies, majorly in December and February of respective years (pronounced in 2021 and 2022). For instance, the hottest temperature was recorded on 15/12/2022, 3/12/2021, and 17/10/2020. The dates of the anomalous peaks were used as ground truth points, which were further investigated and verified by other anomaly detection techniques discussed in the subsequent sections.

Figure 5: Daily average pixel temperatures and outliers identified using visualization

We use descriptive statistics to qualitatively analyze and verify the ground-truthed anomalies of the temporal plots created using the Turkey method. The percentile plot in Figure 6 (A) of average temperature above the background of ROI1 (ΔT) reveals fifteen outliers (anomalies). The majority of the anomalies were observed in February, March, and April of the respective years. On the other hand, the percentile plot for ROI2 (ΔT1) reveals five anomalies, most pronounced in February and October of the corresponding years.

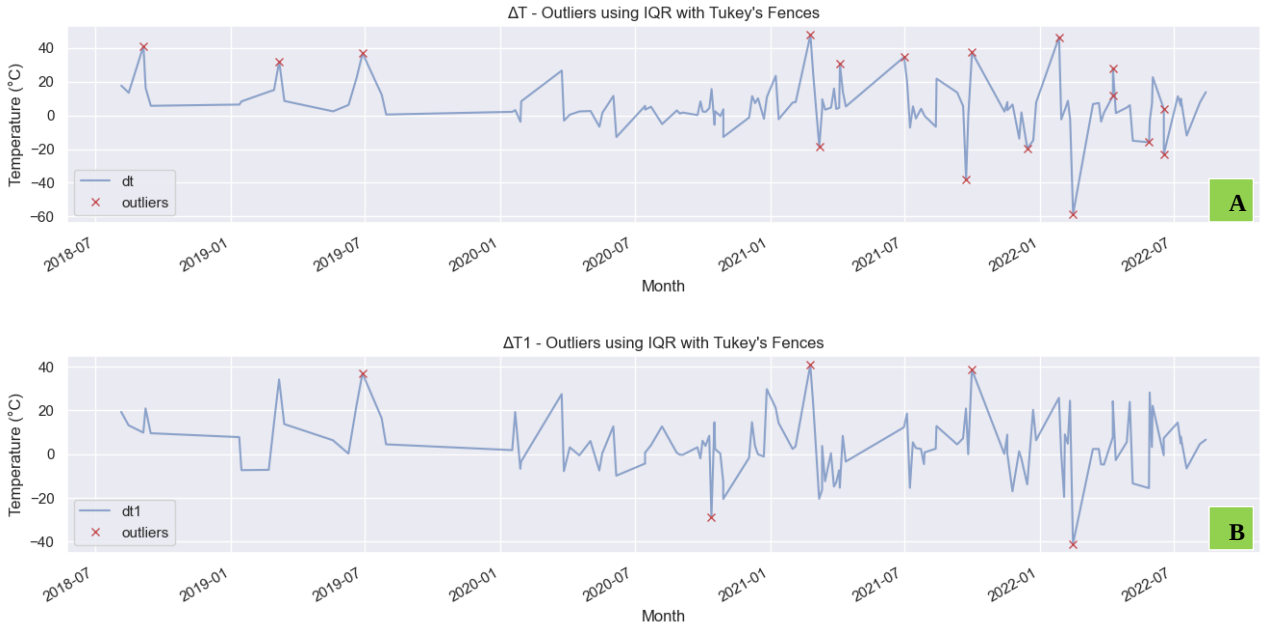

Figure 6: Percentile plots of daily average pixel temperatures above background (ΔT) identified using the Tukey fences method

The seasonal anomaly temporal plots of temperature above the background (Figure 7) agree with the findings of manual flagging and the Turkey method. The grouped pixel temperature values show variation in the respective months. The highest pixel temperature values (extreme maxima) are recorded in the hot and dry months of February. On the contrary, the lowest pixel temperature values (minima) were observed in the wet season (April), and cold periods (July and August).

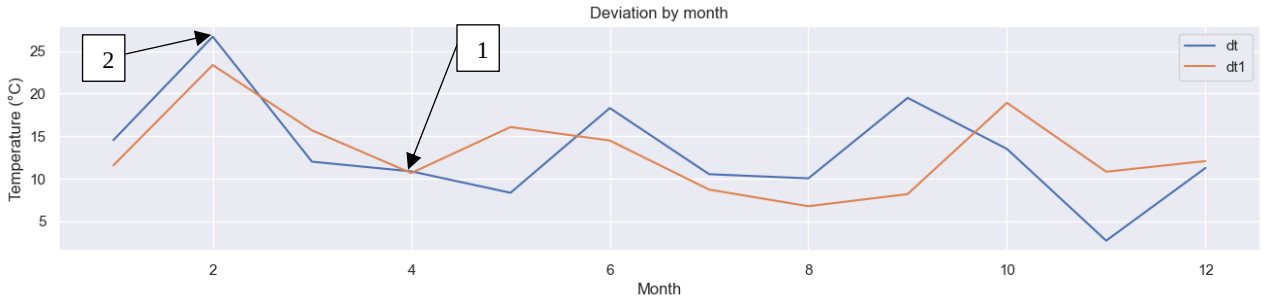

Figure 7: Temporal plots of seasonal average pixel temperature above background plots (grouped monthly). The extreme events (outliers) are annotated by (1):-the highest pixel temperature values and (2):-the low-temperature pixel temperature of respective months.

We observe a similar trend in the findings of the standardization (z-score) method, shown in Figure 8. In this technique, most exceptional events (characterized by standard deviation scores of +2.0 or higher) are observed in February (15/02/2022), March (12/3/2021, 24/3/2020), and October (17/10/2020). Conversely, the extreme events (standard deviation scores of +1.6-2.0) are mainly observed in February (28/2/2021), March (12&7/3/2021), and October (1st &13/10/2021).

Figure 8: Temporal plots of (A &B): standardized temperature profiles showing periods with exceptional extreme anomalous temperature values (scores +2.5), and (C, D): periods with extreme temperature values (standard deviation scores of +1.6-2.0) The red crosses denote outlier events.

Over the same period, the meteorological data (i.e., air temperature data), exhibited exceptional abrupt outlier signals in March (31/03/2019), and April (30/04/2019). The extreme events were mainly identified in March, April, July (31/072022), and August (31/08/2018), as shown in Figure 9. Similarly, the temporal plots of meteorological data, calculated using the seasonal method (Figure 9 C), reveal outliers in February, April, June, and November. These peaks mainly occur during the hottest months (i.e., January to March), cold (June-September), and rainy (April-June, and November) periods.

Figure 9: Temporal plots of average daily temperatures calculated using (A & B): the z-score method, and C: seasonal anomaly detection method.

The statistical comparison between pixel temperatures of the different ROIs vs monthly average meteorological data (49 measurements) is displayed in a corrplot (Figure 10). It is evident from the graph that, the temperature above background (dt, dt1) have a weak positive correlation with air temperature, with a coefficient of determination (r^2), of 0.15 and 0.13, respectively. On the other hand, there is a negative correlation between pixel temperature above background and surface moisture content. The coefficients of determination (r^2) of the moisture content vs pixel temperature above above background (dt, dt1) are -0.19, and -0.07, respectively. The rainfall data were excluded from the study due to the insufficiency of a temporally coherent data over the same period.

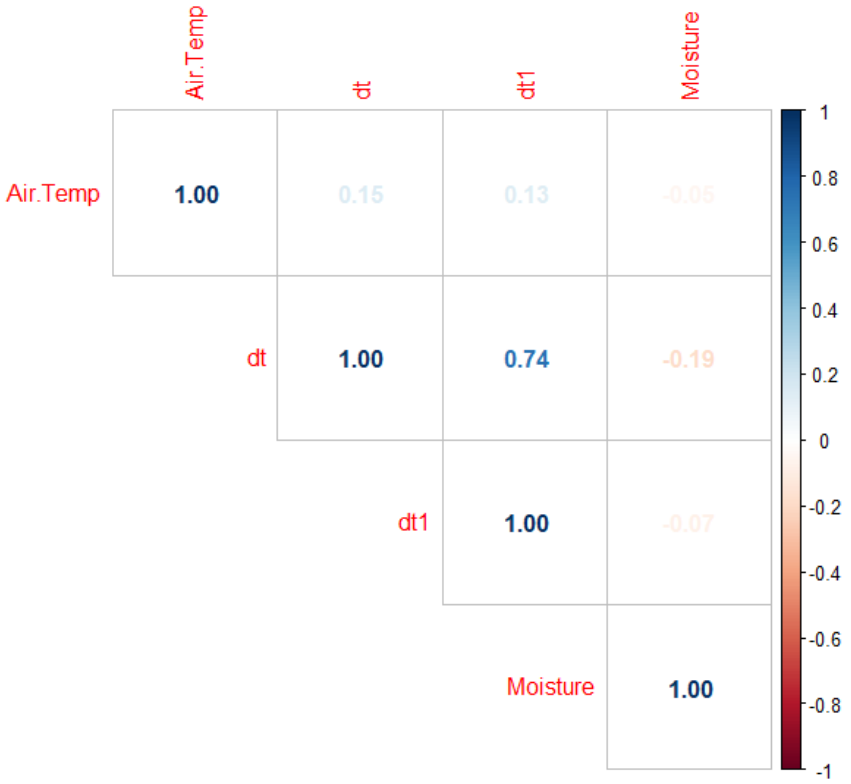

Figure 10: Correlation matrix diagram showing the relationship between air temperature, moisture content and the monthly average pixel temperatures of the various Regions of Interest (ROIs) of the area of focus (49 data points used)

4.4 Discussion

3.1. Correction of atmospheric artifacts on the TIR imagery

It is established that the cloud cover majorly affects the pixel temperatures recorded by ECOSTRESS imageries. These effects were majorly found in the forested areas (South East of the study area), and the northern regions adjacent to the water bodies (e.g., Lake Elementaita). As such, a masking algorithm was used to clip/mask anomalous pixels using quantiles, where pixel temperature values above 98% and below 2% quantiles bracket were flagged out. Most of these cloudy pixels had characteristically low or high kinetic temperature values, deliberated as false positive values, hence unstainable for subsequent analysis. However, the algorithm could not wholly flag out pixels in the South Eastern parts of the study area. These are highly rugged areas, as such speculated to have been affected by other artifacts such as terrain (aspect, slope), and apparent thermal inertia (Kuenzer & Dech, 2013). Such artifacts affect the differential heating or cooling of the surface, thus not pertinent for correction by the cloud masking algorithm.

3.2. Spatio-temporal distribution of surface temperatures

The time series plots of the daily average distribution of pixel temperature of individual ROIs and the computed temperature above the background (ΔT), exhibit a stationary temporal pattern. This implies that (1)-the mean/standard deviation of the respective values is constant over time. As such, proves the study's hypothesis (H_3), which links such behavior to non-geothermal effects. In addition, there is an incremental (spatiotemporally) distribution of pixel temperature values within the vegetation region of interest (ROI2). Such an increase can be attributed to increased vegetation stress mainly due to enhanced evapotranspiration (Legates et al., 2011), as supported by the declining soil moisture trends, acquired over the same temporal span. Likewise, the pixel temperatures of the non-vegetated regions of interest (ROI1 and RO3) increase with time. Such spatio-temporal behavior has been attributed to thermal expansion due to increased surface heating (due to increased solar irradiance) (Diliberto, 2011), or enhanced heat transfer due to hot gas emission along hydrothermal areas (Chiodini et al., 2016). However, the latter hypothesis was difficult to substantiate, owing to the temporal variation of the pixel temperatures (discussed in the subsequent paragraph, and subsections).

The temporal plots from the pixels selected from three Regions of interest, namely: hydrothermal zones (ROI1), highly faulted regions (ROI2), and vegetation (ROI 3) of the focus area exhibit contrasting spatial thermal patterns. The distribution of pixel temperature values in the hydrothermal zones is higher than in the other regions of interest. In contrast, the temporal plots extracted from the raw cloud-masked time series TIR data of the three ROIs exhibit seasonal pattern. Although the temperature values of ROI2 (forested regions) are lower than the pixels from the other plots, they exhibit similar trends. Such oscilating/cyclic behavior can be attributed to diurnal characteristics, which have been established to be controlled by the existing weather conditions (Kuenzer & Dech, 2013), and solar illumination over a 24 hr period. It is also established that contrary to the study's hypothesis, the surface material (varying thermal inertia) did not affect the overall trends of the temporal plots. This can be attributed to the coarse resolution (temporal and spatial); where minute details of the thermal behavior of the materials in the specific ROIs are not captured. Other factors such as pixel mixing (Lillesand et al., 2015), where the prevailing land cover produces contrasting thermal values, are postulated to contribute to the temporal plots'cyclic behavior.

3.3. Anomaly detection results

Different theories have been proposed to explain the possible physical link between the presence of thermal anomalies and hydrothermal unrest or geothermal activities as a proxy for geothermal heat flux (Pavlidou et al., 2016). The main cause includes the expulsion of warm gases and/ or liquids from stressed rock (Chiodini et al., 2016). However, this is a complicated affair, and in most cases, daily or seasonal weather patterns obscure crucial patterns related to thermal events (Kuenzer & Dech, 2013). In this study, we set out to prove/disapprove the aforementioned theory, using four anomaly detection methods namely: manual (visualization) flagging, the Turkey fences method, the seasonal method, and the z-score score technique. Besides, the study aimed to assess whether the detected anomalies could be used to isolate thermally anomalous episodes (often linked to fumarolic activities). Most of the flagged thermal anomalies (as per the findings of all methods) were short-term abrupt signals that depended on the prevailing weather pattern. For instance, positive anomalies (high pixel temperature values) were observed in February, corresponding to hot and dry local climatic conditions. Similarly, negative anomalies (low pixel temperature values) were observed in April. This period relates to periods with intense rainfall content, which is expected to lower the recorded surface temperature. Furthermore, such periods links with enhanced cloud cover and cloud shadows, which obscure transient anomalies (Kuenzer & Dech, 2013), leading to decrease in pixel temperature in areas influenced by these artifact. This was confirmed by the air temperature anomalies, which were peaks in the same month. In addition, seasonal anomalies are observed during dry to wet seasonal transition (commonly occurs between March and April).

Additionally, the outlier signals of the ECOSTRESS pixel temperature plots exhibit a similar quantitative relationship with anomalous signals derived from the meteorological data (air temperature, and soil moisture). It is established that the outlier peaks coincide with those of the air temperature plots, and inversely with the surface moisture plots. This proves the study's hypothesis, which links such anomalous peaks to short-term weather-related events.

It is also observed that there exists a weak (negative and positive) correlation between the pixel temperature values, air temperature and the soil moisture, which falls below study's set hypothesis' threshold of 0.6. This incongruency can attributed to nature of the comparable datasets. Most of the meteorlogical datasets were resampled to attain coherence with the pixel temperature data. This reduced the sample size, which is postulated to have affected the resulting determination of coefficient values.

5 CONCLUSIONS

Satellite-based TIR data is a useful, and promising tool for the detection and monitoring of the thermal behavior of surface conditions in geothermal areas/fields. The advent of higher temporal and spatial resolution spaceborne sensors (e.g., the Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) enhances such prospecting studies. Therefore, the following conclusions were drawn in line with the study's objectives:

1. Cloud cover is the main artifact affecting the recorded pixel temperature values in the study area. These effects are specifically pronounced in heavily vegetated regions, such as the Eburru forest (SE part of the study area). The cloud masking algorithm is an effective tool to filter out these artifacts. However, in this study, other factors such as terrain (slope, and aspect) hindered the effective flagging of these exogenous events.
2. The amplitudes of the background removed time series plots show stationarity over time, an indication that minimal changes have occurred in geothermal hotspot zones. The findings of this study also agree with the previous thermal monitoring studies that have suggested that localized weather patterns affect the pattern of time series data acquired by thermal infrared data. The study qualitatively and quantitatively correlated the time series plots derived from the surface kinetic temperature of ECOSTRESS imagery, and meteorological data. The anomalous signals of the two datasets were established to exhibit empirical similarities.
3. It is possible to flag anomalous signals generated from satellite-based TIR data. However, such signals are mostly sporadic, short-term, weather-induced (false positive), and indirectly linked to fumarole or other hydrothermal activities. The techniques are however promising and could be used for long-term surface temperature monitoring in other geothermal fields.

5.1 Recommendations

The following recommendations were made in line with the study's findings:

1. Apply image correction techniques e.g., (topographic (aspect, slope) correction, in addition to the pre-processing processes applied in this study, to further isolate the extraneous influences that contribute to false positive surface temperature values. In addition, future studies should incorporate hypertemporal TIR data, preferably the stationary thermal cameras in specific hotspot areas, for precise monitoring of short-term fluctuations.
2. Increase the number of training samples for the extraction of pixels with varying material properties, for detailed analysis, and comparison. In addition, automate the ROI creation process, preferably using supervised or unsupervised deep learning techniques.
3. Incorporate the other subsurface geoscientific datasets (e.g., Pressure/Temperature, fumarole chemistry), and compare and evaluate their temporal development. This will be instrumental in linking surface data with physical processes attributed to activities in geothermal systems.

REFERENCES

Atwan, A. T. (2022). Time series analysis with python. In R. Ravikumar, R., Shaikh, T., Moharir, S., Limbachiya (Ed.), *Data Analytics in Cognitive Linguistics* (1st ed.). Packt Publishing Ltd. https://doi.org/10.1515/9783110687279-003

Baroncini, F., Castelli, F., Caparrini, F., & Ruffo, S. (2008). A dynamic cloud masking and filtering algorithm for MSG retrieval of

land surface temperature. *International Journal of Remote Sensing*, *29*(12), 3365–3382. https://doi.org/10.1080/01431160701469032

Chiodini, G., Chiodini, G., Avino, R., Caliro, S., & Minopoli, C. (2011). Temperature and pressure gas geoindicators at the Solfatara fumaroles (Campi Flegrei). *Annals of Geophysics*, *54*(2). https://doi.org/10.4401/ag-5002

Clarke M. C. G., Woodhall D. G., Allen, D., Darling, G. (1990). *Geological, Volcanological and Hydrogeological controls on the occurrence of geothermal activity in the area surrounding Lake Naivasha, Kenya*.

Coolbaugh, M. F., Kratt, C., Fallacaro, A., Calvin, W. M., & Taranik, J. V. (2007). Detection of geothermal anomalies using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) thermal infrared images at Bradys Hot Springs, Nevada, USA. *Remote Sensing of Environment*, *106*(3), 350–359. https://doi.org/10.1016/j.rse.2006.09.001

Dech, S. W., Tungalagsaikhan, P., Preusser, C., & Meisner, R. E. (1998). Operational value-adding to AVHRR data over Europe: Methods, results, and prospects. *Aerospace Science and Technology*, *2*(5), 335–346. https://doi.org/10.1016/S1270-9638(98)80009-6

Diliberto, I. S. (2011). Long-term variations of fumarole temperatures on Vulcano Island (Italy). *Annals of Geophysics*, *54*(2), 175–185. https://doi.org/10.4401/ag-5183

Haselwimmer, C., Prakash, A., & Holdmann, G. (2013). Quantifying the heat flux and outflow rate of hot springs using airborne thermal imagery: Case study from Pilgrim Hot Springs, Alaska. *Remote Sensing of Environment*, *136*, 37–46. https://doi.org/10.1016/j.rse.2013.04.008

Henderson, S. T., Pritchard, M. E., Cooper, J. R., & Aoki, Y. (2019). Remotely Sensed Deformation and Thermal Anomalies at Mount Pagan, Mariana Islands. *Frontiers in Earth Science*, *7*(September), 1–14. https://doi.org/10.3389/feart.2019.00238

Jimenez-Munoz, J., Mattar, C., Sobrino, J., Malhi, Y., 2015. (2015). A database for the monitoring of thermal anomalies over the Amazon forest and adjacent intertropical oceans. *Scientific Data*, *2*, 1–9. https://doi.org/10.1038/sdata.2015.24

Kuenzer, C., & Dech, S. (2013). Remote Sensing and Digital Image Processing Thermal Infrared Remote Sensing. In F. D. Van Der Meer (Ed.), *Remote Sensing and Digital Image Processing*. Springer Science+Business. https://doi.org/DOI 10.1007/978-94-007-6639-6_1,

Legates, D. R,. Mahmood, R., Levia, D. F., DeLiberty, T. L., Quiring, S. M., Houser, C., Nelson, F. E. (2011). Soil moisture: a central and unifying theme in physical geography. *Prog Phys Geogr*, *35*(1), 65–68.

Lillesand, M, T., Kiefer, W, R., Chipman, W, J. (2015). *Remote Sensing and Image Analysis* (7th ed.). Wiley & Sons, Inc.

Malvina, S., Rabuffi, F., Pisciotta, A., Musacchio, M., Diliberto, I. S., Spinetti, C., Lombardo, V., Colini, L., & Buongiorno, M. F. (2019). Analysis of thermal anomalies in volcanic areas using multiscale and multitemporal monitoring: Vulcano Island test case. *Remote Sensing*, *11*(2). https://doi.org/10.3390/rs11020134

Mia, M. B., Bromley, C. J., & Fujimitsu, Y. (2012). Monitoring heat flux using Landsat TM/ETM+ thermal infrared data - A case study at Karapiti ('Craters of the Moon') thermal area, New Zealand. *Journal of Volcanology and Geothermal Research*, *235–236*, 1–10. https://doi.org/10.1016/j.jvolgeores.2012.05.005

Murphy, S. W., Filho, C. R. de S., & Oppenheimer, C. (2011). Monitoring volcanic thermal anomalies from space: Size matters. *Journal of Volcanology and Geothermal Research*, *203*(1–2), 48–61. https://doi.org/10.1016/j.jvolgeores.2011.04.008

Mwarania, F. M. (2014). *Reservoir Evaluation and Modelling of the Eburru Geothermal System , Kenya* (U. N. University (ed.); Issue December).

Omenda, P. A. (2012). Geothermal Development in Kenya : A Country Update - 2012. *ARGeo-C4*, *Fourth African Geothermal Conference*, 89–93.

Pastor, F., Valiente, J. A., & Khodayar, S. (2020). A warming Mediterranean: 38 years of increasing sea surface temperature. *Remote Sensing*, *12*(17), 1–16. https://doi.org/10.3390/RS12172687

Pavlidou, E., van der Meijde, M., van der Werff, H., & Hecker, C. (2016). Finding a needle by removing the haystack: A spatio-temporal normalization method for geophysical data. *Computers and Geosciences*, *90*, 78–86. https://doi.org/10.1016/j.cageo.2016.02.016

Saraf, A. K., Rawat, V., Choudhury, S., Dasgupta, S., & Das, J. (2009). Advances in understanding of the mechanism for generation of earthquake thermal precursors detected by satellites. *International Journal of Applied Earth Observation and Geoinformation*, *11*(6), 373–379. https://doi.org/10.1016/j.jag.2009.07.003

Shi, J., & Hu, C. (2021). Evaluation of ecostress thermal data over south florida estuaries. *Sensors*, *21*(13). https://doi.org/10.3390/s21134341

Vaughan, R. G., Keszthelyi, L. P., Lowenstern, J. B., Jaworowski, C., & Heasler, H. (2012). Use of ASTER and MODIS thermal infrared data to quantify heat flow and hydrothermal change at Yellowstone National Park. *Journal of Volcanology and Geothermal Research*. https://doi.org/10.1016/j.jvolgeores.2012.04.022

Vilardo, G., Sansivero, F., Chiodini, G. (2016). Long-term TIR imagery processing for spatiotemporal monitoring of surface thermal features in volcanic environment: A case study in the Campi Flegrei (Southern Italy). *Journal of Geophysical Research: Solid Earth*, *1*, 3782–3803. https://doi.org/10.1002/2015JB012608

Appendix 1: python script for Cloud masking of ecostress images

```
# STEP 1: Import the necessary libraries

from os.path import join, basename,splitext

import numpy as np

from datetime import datetime

import rioxarray

import hvplot.xarray

import xarray

import rasterio

# STEP 2: Load and visualize a specific ECOSTRESS image

input_dir=r'E:\Research\ECOSTRESS\Jan_Dec_2022\Clean images' # Specify the input directory

ST_filename=join(input_dir,'ECO2LSTE.001_SDS_LST_.tif') # Read an image from input directory

datetime_UTC=datetime.strptime(splitext(basename(ST_filename))[0].split('_')[-2][3:],'%Y%j%H%M%S') # Convert a specific column of the image dataframe into datetime

ST=rioxarray.open_rasterio(ST_filename).squeeze('band',drop=True) # Open the raster image

ST.data=np.where(ST.data==0,np.nan,ST.data*0.02)-273.15 # Convert the pixel values to Degree Celcius

ST.rio.reproject('EPSG:4326').hvplot.image(cmap='jet',alpha=0.7,title=f'Surface Temperature for Eburru Geothermal Prospect {datetime_UTC:%Y-%m-%d %H:%M:%S}UTC'

,width=720,height=720) # Transform the image's coordinate system and visualize the raw ECOSTRESS image (with dimensions of choice)

# STEP 3:  Load and visualize a specific ECOSTRESS cloud masked image

input1=r' \Cloudss'# Specify the input directory

Cloud_filename=join(input1,'ECO2CLD.001_SDS_CloudMask_.tif') # Read a cloud mask image from the input directory

datetime_UTC=datetime.strptime(splitext(basename(Cloud_filename))[0].split('_')[-2][3:],'%Y%j%H%M%S') # Convert a specific column of the image dataframe into datetime

Cloud=rioxarray.open_rasterio(Cloud_filename).squeeze('band',drop=True).rio.reproject_match(ST) # Match the cloud masked file's coordinate system to the ECOSTRESS image

Cloud.data=(Cloud.data>>2) & 1 # Specify a condition

Cloud.rio.reproject('EPSG:4326').hvplot.image(cmap='jet',alpha=0.7,title=f'Surface temperature_Eburru Geothermal Prospect {datetime_UTC:%Y-%m-%d %H:%M:%S} Binary'

,width=720,height=720)  # Transform the image's coordinate system and visualize the raw ECOSTRESS image (with dimensions of choice)
```

Omwenga and Bonyo.

```
# STEP 4: Merge the images to produce a cloud masked ECOSTRESS image

ST.data=np.where(Cloud.data,np.nan,ST.data)

ST.rio.reproject('EPSG:4326').hvplot.image(cmap='jet',alpha=0.7,title=f'Surface Temperature for Eburru Geothermal Prospect {datetime_UTC:%Y-%m-%d %H:%M:%S} Masked'

,width=720,height=720)

ST.rio.to_raster(r'E:\Research\ECOSTRESS\Output files\220127_masked.tif') # Write the cloud masked tif file into a directory of choice
```

Appendix 2: A python script for extraction of time series plots from bulk images

```
Step 1: # Load necessary libraries

import geopandas as gpd

import rasterio

import rasterstats

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

# Run a loop over the files, and read info of the files

i=0

for files in os.listdir(r'E:\...\ECOSTRESSFOLDER '):    # Run a loop over a specific directory

   if files[-4:]=='.tif': # Only read files ending with .tif

      i=i+1

      dataset=rasterio.open(r'E:\...\ECOSTRESSFOLDER '+'\\'+files)

      x,y=(36.282427,-0.540210) # Create x and y coordinates

      row,col=dataset.index(x,y) #Create new variable called row and column to retrieve the coordinates

      data_array=dataset.read(1) # Create a new variable called data array

      # Copy the Date into the 'Date' column into the table during each iteration

      table['Date'].loc[i]=files[:-4]

        # Copy the Date into the 'Rainfall' column into the table during each iteration

      table['Rainfall_mm'].loc[i]=data_array[int(row),int(col)]

      # Export table into csv

      table.to_csv((r'E:\...\ECOSTRESSFOLDER \ECOSTRESS_TIME_SERIES.csv')
```

Appendix 3: Stationarity test-python script and results

```
# Augmented Dickey Fuller (ADF) test

# step 1: Load .csv data into a pandas dataframe, and read a column of interest (‘ROI1’)

series=pd.read_csv(r"…..\ECOSTRESS_2018-2022_mod.csv",index_col='Month', parse_dates=True)

series= series['ROI1']      # Select column of interest

# step-2: Create an ADF variable from the selected column
```

```
series_ADF=series.values  # ADF variable
# Create another variable using ADF module
ADF_test = adfuller(series_ADF)
#step-3: Loop over the values to test for stationarity
for key, value in ADF_test[4].items():
   print('\t%s: %.3f' % (key, value))
if ADF_test[0] < ADF_test[4]["5%"]:
   print ("Reject Ho - Time Series is Stationary")
else:
   print ("Failed to Reject Ho - Time Series is Non-Stationary")
# Results
1%: -3.485
        5%: -2.885
        10%: -2.579
Reject Ho - Time Series is Stationary
# step 5: Repeat the process for the other ROIs
```