Utilizing Historical Satellite Data for Spatiotemporal Detection of Disaster-related Anomalies

Bienvenido G. Carcellar III[1, 2] and Masahiko Nagai[2]

[1] Department of Geodetic Engineering, University of the Philippines – Diliman, Quezon City 1101 Philippines
[2] Yamaguchi University, Center for Research and Application of Satellite Remote Sensing (YUCARS), 2-16-1 Tokiwadai, Ube, Yamaguchi, 755-8611, Japan
Corresponding author: bgcarcellar@up.edu.ph

Abstract – *With the increasing accessibility of historical satellite datasets and the emergence of satellites with higher temporal resolution, the geospatial community faces the challenge of effectively utilizing this wealth of past information. This research aims to leverage this high temporal resolution satellite dataset to uncover long-term patterns in historical satellite data, particularly in urban areas impacted by disasters. The study focuses on baseline modelling and anomaly detection of VIIRS SNPP day/night bands VNP46A1 and VNP46A2. Preprocessing and transformation of satellite images into a pixel-based time series dataset were performed, followed by progressive harmonic modelling using one-year lag data and three-month progression parameters to generate baseline models. Subsequently, anomaly detection was conducted using fbprophet, with an optimal value of 120-day time lag used to calculate the interval width. The algorithm successfully detected anomalous values, which were then mapped and correlated with significant events occurring in the selected study areas, including severe flooding and earthquakes. The results demonstrated the algorithm's capability to extract valuable information from large historical satellite dataset and highlight its potential for real-world applications.*

Keywords: Nighttime Light; Disaster Mitigation; Anomaly Detection; Time Series Analysis

I. INTRODUCTION

For the past years, various natural and artificial disasters around the globe affected many people's lives. Disasters don't choose their victims; they may hit either rural or urban areas, and they may take the lives of people regardless of status. However, disasters in a built environment pay more significant risks [1]. Effects such as human, financial, and economic losses are higher, especially for developing countries, since most of the population resides and economic activities happen in urban areas [2]. As such, monitoring disaster-prone urban areas is essential in mitigating damage to these areas.

Several vital cities are identified as disaster-prone, especially along the Pacific Ring of Fire. These cities usually expect several natural disasters in a year. For the said cities, identifying insights from previous disasters is a helpful approach to better preparing such cities for possible future disasters [3]. Further examining previous disasters by finding considerable changes in environmental indicators before, during, and after the disaster may lead to valuable findings in monitoring these locations. Past information can be analyzed to gain insight into what happened, which leads to a proper assessment of the damage and recovery of the area for future planning [4].

One common source of information helpful in understanding the effects of disasters on the physical environment is satellite images. During and after the disaster, several satellite images are available to provide information for determining the impact of the disaster on the area [5]. In this era of information, satellite datasets provide us with a wide array of historical snapshots of the world [6]. The vast historical library of satellite data available to analyze is getting wider as technological advancements have made it possible to store records of satellite images. With several years of information from satellite data, it has become a challenge to utilize such a considerable amount of historical satellite data [7]. Historical data analysis plays a huge role in such use cases, giving us valuable past information.

Several satellite datasets have been used for various environmental applications since the early 2000s. Particularly for an urban environment, the Suomi National Polar-orbiting Partnership (SNPP) Visible Infrared Imaging Radiometer Suite (VIIRS) Day-Night Band provides information for nightlight radiance, which is often associated with economic activity in urban areas [8]. VIIRS Day-Night band can be used to view the sudden power cuts associated with disasters [9, 10]. These power cuts' extent and area of effect can be extracted from these images [11]. The temporal resolution of the VIIRS Day-Night band is daily [8]. Such data can be appropriate for in-depth time series analysis, providing ample data for forming baseline historical data.

One common analysis for historical or time series datasets is anomaly detection. There is a wide range of applications in which anomaly detection can be utilized. It is often used in cybersecurity, finance, and communications [12]. A critical aspect of anomaly detection involves massive amounts of information and finding outliers from these datasets. Because of this nature of anomaly detection, numerous studies provide innovations in finding anomalous information from the data through data mining techniques [13]. This is true for mostly continuously streaming information sources such as finance, health records, and surveillance. This setup is also applicable in the case of historical satellite datasets, which is the aim of this research. Such historical datasets can be viewed as streaming information in which the satellites provide access and surveillance to the subject area, the Earth. It is important to note that the enormous historical data problem for anomaly detection persists until today, and it is considered a curse of dimensionality, making it hard to generate accurate and high-performing anomaly detection models [14].

Because the geospatial community recently began to realize the increasing amount of historical satellite data, there was also increasing interest in anomaly detection studies in satellite imageries. Early anomaly detection processes in satellite data, which were very common in the early 2000s, were change detection studies [15]. Few studies focus on the method of anomaly detection for continuous time series and expanding the time scale to accommodate more historical satellite images. Several other satellite time series anomaly detection models were developed, primarily relying on historical data statistical patterns [16, 17, 18, 19, 20, 21]. These models can be further divided into supervised and unsupervised anomaly detection methods.

Supervised models utilize a labeled training and test dataset to investigate time series

patterns. A common ground to such models is calculating the probability of an area being of a particular land cover class based on historical data [17, 18, 21]. Such probabilistic models require vast amounts of training datasets that are normal or non-anomaly, making monitoring and performing difficult.

Most anomaly detection studies on satellite datasets utilize unsupervised learning algorithms to determine and label anomalous events [22]. This approach is particularly suited when the subject being analyzed doesn't change throughout multiple captures of satellite images, making it challenging to pinpoint and label the datasets. One such case of studying continuous datasets is by Zhou and Tang [15], in which they developed an anomaly detection model. This model utilized the Z-score of Season-Trend model Residuals (ZSTR). It successfully detected regions with severe flooding. Other studies aiming to identify anomalies through unsupervised classification used several statistical concepts such as the Gaussian Process for probabilistic forecasting [20], parametric decomposition of time series to linear trend, seasonality, and bias components [16], and parametric model based on triply modulated cosine models on a sliding window. In line with this, the study adopted an unsupervised methodology, employing both unsupervised anomaly detection algorithms and harmonic models. The study aims to utilize concepts similar to Verbesselt, Zeiles, and Herold [16] by using harmonic models in parametrizing the time series model. This study also utilized the parametrization concept from Chandola and Vatsavai [20] in decomposing the dataset into its trend, seasonality, and bias components by fitting a harmonic model and identifying its harmonic parameters. Finally, a moving window concept used by Verbesselt, Zeiles, and Herold [16] was also used in this study. Both algorithms used in the study mainly relied on identifying existing patterns within the dataset.

This study highlighted the importance of utilizing a library of historical satellite data for creating baseline models. The research used past data and identified severe disasters in the historical dataset as anomalies, especially in urban areas. A baseline model was generated using harmonic modelling, which described the general trend of nightlight radiance in the study area. After which, an anomaly detection model using fbprophet was created that identified anomalies in a time series of satellite images via pixel-wise time series analysis. Further, the identified anomalies in historical satellite datasets were then correlated to disasters.

II. MATERIALS AND METHODS

2.1 Research Concept
The research concept relied on finding anomalous values of observed information in the satellite data compared to the generated baseline. The research centered on historical satellite data and the relevant disasters in an area. The concept of the research was to tie in the two main ideas through mathematical modelling of the historical satellite data and anomaly detection within the said data, as shown in Figure 1. In this study, the anomalies detected were correlated as relevant disasters with a high probability of correlation with the abnormal values of radiances.
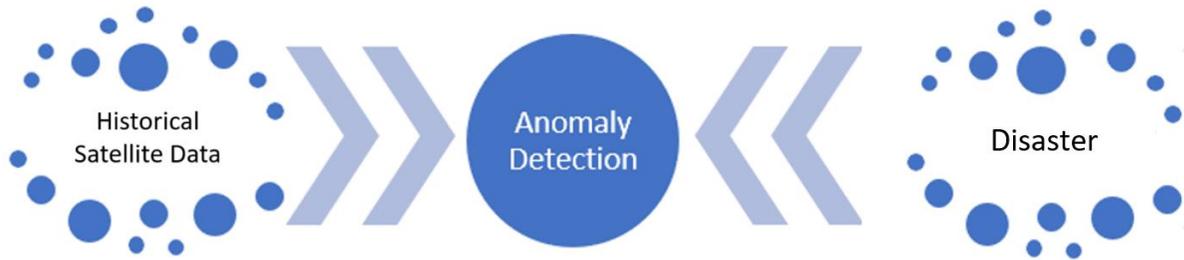
Figure 1. The research concept. The arrows pointing toward anomaly detection represent how both historical satellite datasets and relevant disasters can be correlated through anomaly detection.

2.2 Research Framework

The research framework is shown in Figure 2. It was divided into three phases: satellite data preprocessing, baseline modelling, and anomaly detection. The specifics of the three main phases are discussed in the following subsections.
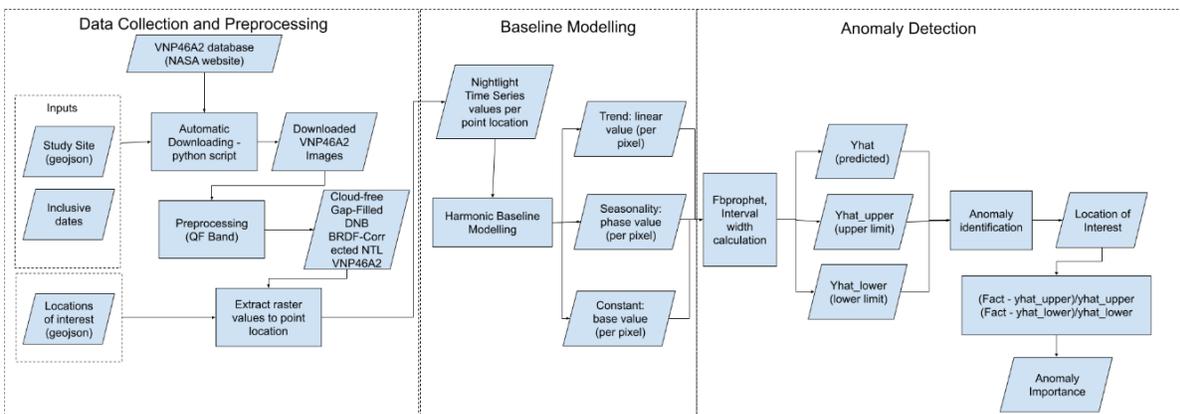

Figure 2. The research framework is divided into three phases: Data Collection and Preprocessing, Baseline Modelling, and Anomaly Detection.

The first phase initially managed and organized the satellite dataset from their sources. As working with historical datasets was quite challenging, it was essential to have a scheme to organize such datasets for easier access to information for later use. This phase focused on preprocessing these satellite images and managing the images to a more accessible online platform. In this research, the satellite images used were VIIRS/NPP Nighlight radiance products VNP46A1 and VNP46A2 from NASA's Goddard Space Flight Center (GSFC) Black Marble nighttime lights product suite. The output of this phase was the preprocessed images readily available for a pixel-based time series analysis. The second phase of the research framework was the baseline modelling. This part utilized the output of the first phase and transformed it into pixel-based time series data. The final product of this phase was a baseline model for each pixel of the study area. For the last phase, the original time series dataset was subjected to an anomaly detection algorithm to determine which values the model determined

to be anomalous. The anomaly detection utilized several parameters, including those describing the baseline model output from the second phase. The anomaly values and the corresponding dates on which they were recorded were considered to determine the extent and the dates of the anomaly values. These were then correlated with the occurrence of several relevant disasters.

2.3 Study Area

Two study areas were chosen to detect the anomaly of night light. Various disasters historically damaged these areas during the inclusive dates of VIIRS SNPP day/night bands from 2012 to the present. Figure 3 shows the location of the study areas of the research.
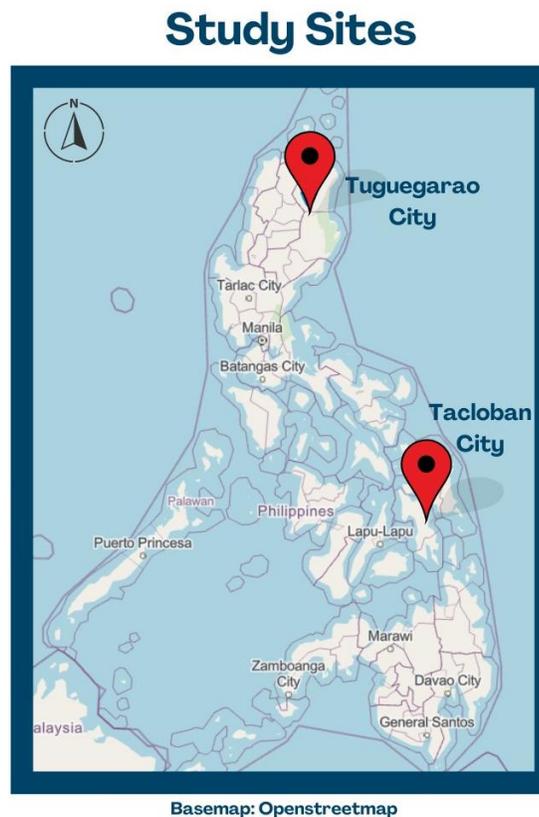
Study Sites

Figure 3. The study sites of the research – Tuguegarao City and Tacloban City, Philippines

The first site, Tacloban City, Philippines, is one area often hit by disasters. From 2012 to 2013 alone, the city was hit by three disasters, which were so devastating that it affected many lives. The first disaster was a 7.6 magnitude earthquake that hit the eastern Philippines last August 31, 2012 [23]. Reported power outages due to the quake were received. The next disaster happened a few months after the first, which was a tropical storm. On Christmas day, the tropical storm Wukong hit the island of Samar and ravaged several cities, including Tacloban [24]. Flooding in the area was experienced as heavy to intense rain hit the city. Like the first disaster a few months before, power outages were also experienced due to the strong winds and floods. The most devastating disaster happened about a year after the said storm.

Super typhoon Haiyan, one of the strongest typhoons encountered in the world, hit Tacloban city [25]. The said typhoon destroyed several parts of the city as several houses were destroyed and many lives perished.

As this was the case for the city, the condition of the multiple disasters in the area made it an interesting site for anomaly detection. Since these disasters caused power outages in the place, using the Nightlight radiance dataset was suitable.

The second disaster case examined in this study was Typhoon Vamco, one of the strongest typhoons to hit the Philippines last November 11, 2020. Several cities were flooded, and even in Metro Manila, heavy rains hit the area, often compared to the rainfall of Tropical Storm Ketsana, which brought 445 mm of rain in 24 hours [26]. The study area chosen was Tuguegarao, which was heavily flooded.

2.4 Datasets

There were two types of datasets involved in this research: satellite images and the relevant disasters. The latter was collected and researched through news clippings and online articles about the specific disasters in the study area for each case study. As for the former, the satellite datasets were the primary sources of scientific information. This research's two main satellite data sources were NASA's GES DISC and GSFC Black Marble data products. The specifics of the datasets and the procedures performed to the satellite datasets were detailed in the succeeding subsections.

2.4.1 Data Sources

VNP46A1 and VNP46A2 datasets were intended to be chosen as the data products for two reasons. The first reason was because of its free-to-access characteristic. Downloading the whole archive of satellite images for both products was not a hindrance since it is freely accessible. The second reason was because of its temporality. The data products used for this research were all released daily. This means that new satellite observations can be incorporated into the time series model for each day. This set of characteristics for both datasets sufficed the need for the daily time series analysis to work.

The primary dataset used for this research was from VIIRS/NPP. The SNPP VIIRS supports a Day-Night Band [11] sensor. This sensor measures nightlight from visible and near-infrared (NIR) light. The sensor is ultra-sensitive, especially in lowlight conditions, which produces high-quality nightlight radiance measurements [11].

There are two main VIIRS DNB-based data products. The first one is the VNP46A1, which has a formal name of VIIRS/NPP Daily Gridded Day Night Band. VNP46A1 provides top-of-atmosphere at-sensor nighttime radiance measurement. The second product, VIIRS/NPP Gap-Filled Lunar BRDF-Adjusted Nighttime Lights or VNP46A2, provides much better and more accurate nightlight radiance values than VNP46A1 [11]. VNP46A2 provides a corrected nightlight radiance value, removing the moon illumination fraction and applying atmospheric corrections. Both products are supplied as a 10-degree x 10-degree grid with a spatial resolution of 15 arc seconds. Both also provide radiance measurement of nightlight, but VNP46A2 delivers a much more accurate value of nightlight radiance than VNP46A1 [11].

Both are available to download from the GSFC archive and distribution system and are open to the public [8].

For VNP46A1, the archive downloaded via the online distribution system was from January 19, 2012, up to November 28, 2020. As for VNP46A2, the available dataset in the same distribution system was from January 19, 2019, up to September 20, 2018. The Black Marble team has been constantly updating and uploading the data product to their distribution system daily until it catches up to a same-day upload, such as in VNP46A1. Although both products provided nightlight radiance values, because of the incomplete archive value of VNP46A2, both products were used and considered for this research. The VNP46A1 was explicitly used for more recent cases, such as the disaster case in Tuguegarao, as discussed in Section 3. Meanwhile, for the general case in which nightlight was observed, VNP46A2 was the default dataset used.

2.4.2 Data Collection and Storage

Since the datasets are massive in amount and size, proper data collection was implemented for both datasets. For VIIRS VNP46A1 and VNP46A2, the archive was downloaded remotely from the NASA website using Python code. After gaining access to a unique token, the code stores the downloaded data to the connected cloud server (Google Drive). The token and the downloading of the dataset can be accessed for free through registration. The code was developed using several Python packages such as argparse, os, shutil, csv, JSON, pathlib, sys, re, and glob. It was done using the online jupyter notebook Google Colaboratory. The datasets can be easily stored in the connected Google Drive. This code was used for both VNP46A1 and VNP46A2.

For a single day, the highest total number of satellite images uploaded is 648. The dataset released is divided into 10-degree x 10-degree grids, each containing the nightlight radiance values in a 15-arcsecond spatial resolution. This meant that a single image has 2400 x 2400 pixels, which covers the whole 10 x 10-degree grid. Since many satellite images needed to be downloaded, only selected grids were downloaded to analyze the nightlight radiance. The downloaded archive images for both VNP46A1 and VNP46A2 were dated way back from January 19, 2012, until September 20, 2018, for VNP46A2 and November 28, 2020, for VNP46A1 for a total of approximately 1 TB worth of satellite images. The download time for the historical dataset (worth about nine years) for a single grid was around one week. Regarding its velocity, the datasets are uploaded daily, which is relatively high compared to other freely available satellites.

2.5 Data Pre-processing

After the images were downloaded and stored, a Python script was developed using Google Colab to extract the needed information. The script used several Python packages to open the images and convert them to proper formats. The package h5py was used to open EOS5 Hierarchical Data Format (HDF-EOS5) format for downloaded images. The math and numpy packages were used to transform the data into a matrix for a more straightforward calculation of the image values. The datetime package was used for getting the image dates and properly creating a time label for the images.

For VNP46A1, the band DNB at Sensor Radiance was the main band used in getting the nightlight radiance data. The pixels with clouds were removed using the QF Cloud mask band of the same image. The cloud masking scheme used in this research relied on the cloud mask quality values and confidence indicators. The values selected for cloud mask quality were high and medium, while the poor and low cloud mask qualities were automatically removed. For the cloud mask confidence indicator, the selected values were confident clear, and probably clear. The confident cloud and probably cloud were also removed.

For VNP46A2, the Gap-Filled DNB BRDF-corrected NTL band was utilized for getting the radiance values. The satellite data also contained a cloud masking band to remove pixels with clouds. The same scheme from VNP46A1 was used in operating the cloud mask band. Aside from this, the mandatory quality flag band was also used in further filtering the pixels used in the main algorithm. Pixels that were identified as persistent nighttime lights in the mandatory quality flag were selected to be used for the time series analysis. Other pixels flagged as ephemeral nighttime lights, and outliers were removed further. The values of the band are shown in Table 1.

Table 1 Values of the Mandatory quality flag in VNP46A2 product
and their corresponding meanings

Value	Retrieval quality	Algorithm instance
00	High-quality	Main algorithm - Persistent NTL (e.g., city lights, electrical infrastructure)
01	High-quality	Main algorithm - Ephemeral NTL (e.g., fire, forest fire)
02	Poor-quality	Main algorithm - outlier, potential cloud contamination
255	No retrieval	Fill value

2.6 VNP46A1/VNP46A2 Baseline modelling

The process of baseline modelling for VNP46A1 and VNP46A2 was the same, and this section details the process of getting the baseline model. As mentioned, the globe was divided into 10-degree x 10-degree grids containing 2400 x 2400 pixels. The grids used for analysis were h30v07 for Tuguegarao, Philippines, and h30v08 for Tacloban, Philippines.

A pixel-based time series was generated for each pixel in the image. The values for a single pixel were stacked in a time series, including the corresponding time labels when the image was acquired. This was done using Python script and several packages such as numpy and pandas. The process of baseline modelling is further visualized in Figure 4.
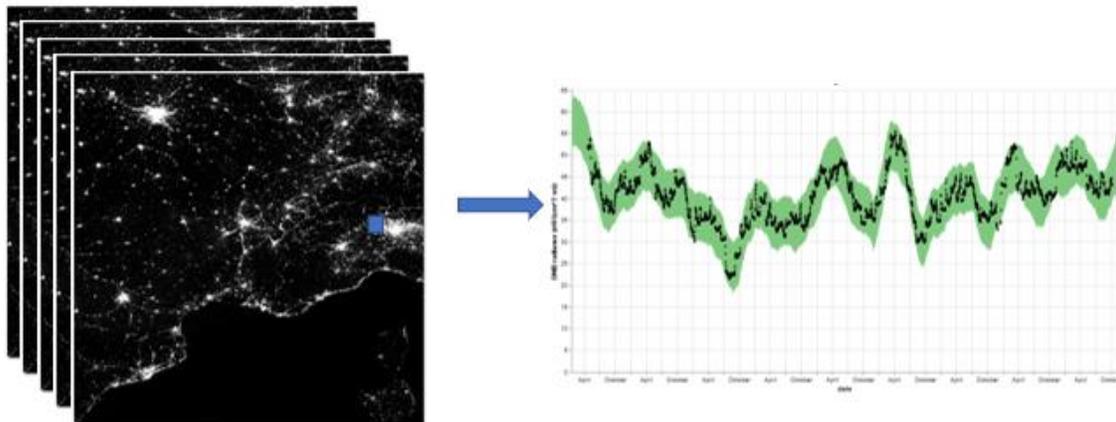
Figure 4. VIIRS Nightlight Baseline Modelling. The blue square in the figure above corresponds to a pixel visualization. Since the model generation is pixel-based, all the archive data were stacked, and a single time series data was generated for each pixel.

For VNP46A1, the datasets stacked were from January 19, 2012, to November 28, 2020. Meanwhile, for VNP46A2, the dates included in the time series were from January 19, 2012, to September 20, 2018. For the case of Tacloban, VNP46A2 was used since this is the default dataset for the inclusive dates. For the case of Tuguegarao, the dataset used was VNP46A1. This is because the dates for the disasters include those beyond September 20, 2018, which was the date limitation for VNP46A2.

Harmonic regression modelling has been utilized and demonstrated as an effective method in remote sensing studies [27, 28]. The mathematics for harmonic regression relies on the assumption that the dataset has periodic and linear patterns that can be described mathematically. The harmonic model can be generally viewed as shown in Equation 1 below by Shumway and Stoffer [29].

$$p_t = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi\omega t) + \beta_3 \sin(2\pi\omega t) + e_t \qquad (Eq.\ 1)$$

The equation provides constant and linear terms of β_0 and $\beta_1 t$. The constant term provides a base value that can be interpreted as a minimum value required for the observed data. The linear term provides a general trend beyond the seasonal trend captured in the two harmonic parameters. These four parameters generally describe the harmonic model. In this research, a variant of the harmonic model is used, and the central equation is still retained. The main idea of progressive harmonic modelling used in this research is about grouping the time series data set into shorter and more nearby time lags to generate a dynamic time series baseline model.

The generated time series for each pixel were also subjected to a 90-day rolling average to fill in predictions to dates without values. Afterwards, the averaged time series was subjected to similar progressive harmonic modelling to generate the model parameters. The harmonic modelling angular frequency was set to one year. Although the model can be developed given a minimum of 6 months ' worth of data, the ideal sample size for such a case is at least a year's

worth for a better model generation from several parameter tuning. Following a one-year lag data, the harmonic modelling process emphasizes recent datasets more by utilizing a three-month progression while giving lesser priority to data values older than a year. Because of the nature of the progressive harmonic modelling, the model parameters vary per three months. This means that for each pixel, the model parameters amplitude, phase, constant term, and linear term have 59 sets of values. The progression starts in 2013 when there was at least one year's data. Similarly, the progression rate in this case is three months, leading to 23 sets of parameter values for VNP46A2 and 29 sets for VNP46A1.

Finally, the same procedure was done to all pixels in the whole image with a size of 2400 x 2400 pixels. After generating all the parameters, these were stored in a Google Drive to access in the third part, anomaly detection.

2.7 Anomaly Detection

For anomaly detection, the time series datasets were subjected to an anomaly detection algorithm of fbprophet. In this algorithm, the model takes in the values of the parameters from the initial baseline model and other parameters. The fbpropeth anomaly detection's main essence is calculating the interval width. The interval width acts as the confidence interval that can be added and subtracted to the baseline model, the part of which can be considered normal and not anomalous. If a value falls beyond the interval width, the value itself is considered anomalous in nature. The list of parameters needed to be determined to solve for the optimal interval width can be seen in Table 2.

Table 2. The list of parameters in calculating the interval width

Interval width parameters	Definition
Time lag	Number of previous observations to be considered relevant
Time lag weight	Time lag multiplier
Seasonal lag	Number of observations in the previous year of the same date to be considered relevant
Seasonal lag weight	Seasonal lag multiplier
Baseline parameters	Constant, linear, and harmonic terms

The calculations of interval width were done per value in the time series. Parameter tuning was done to selected pixels to determine the most appropriate value of time lags. The optimal value for time lag is 120 days. The seasonal lag was for two years in VIIRS Nightlight. The time lag weight carries an exponential value with an exponential rate of 2. The initial weight is set as 1, which is given to the immediate lag, and the weight values decrease exponentially as the observations go farther than the data acquisition date. The seasonal lag weight was employed as a simple linear decay from a weight of 1 to 0.75, 0.5, and 0.25. The baseline parameters aid in calculating the baseline and comparing the actual data to it.

2.8 Time series anomaly event correlation

During the anomaly detection phase, interval widths are generated to identify anomalies in each pixel and determine if they are related to relevant events within a specific timeframe. Comparing the actual values with the interval width makes identifying anomalies in a single-pixel time series easier. For this output, specific pixels were selected based on the visual analysis of optical images in Google Earth Historical Imagery, identifying the areas severely affected by the event.

After choosing the pixel for analysis, a time series chart was generated, which contained the original values and the interval width calculated. The anomalies were identified using the interval widths, and the dates on which the time series values fell as an anomaly were recorded. The recorded dates were then compared with the dates of the relevant events that occurred in the study area. The anomalies were plotted in red for further visualization and were cross-checked with news articles to ensure accuracy.

2.9 Anomaly Mapping

To improve the correlation of the time series anomaly event, we created maps that display the anomalies on different dates. This was done automatically using a Python script we developed in Google Colab. The script identified the dates and values of the anomalies and geotagged them using the pixel location corresponding to the time series. We then converted the collection of anomalies into shapefiles and viewed them in QGIS for mapping. With QGIS, we could map out the anomalies' locations and style the anomaly points' appearance based on their values. Using the time manager plugin in QGIS, we generated single frame maps per day, which were then combined into a single GIF, depending on the cases covered.

III. RESULTS AND DISCUSSION

3.1 Typhoon Haiyan and other disasters in Tacloban and Samar

Figure 5 shows the time series anomaly graph for the event correlation of the chosen pixel. The black points correspond to the actual time series datasets. The green shade is the interval width, in which values that fall below or above the data are anomalous. The larger the red circles are, the higher the deviation of the actual data from the interval width calculated.
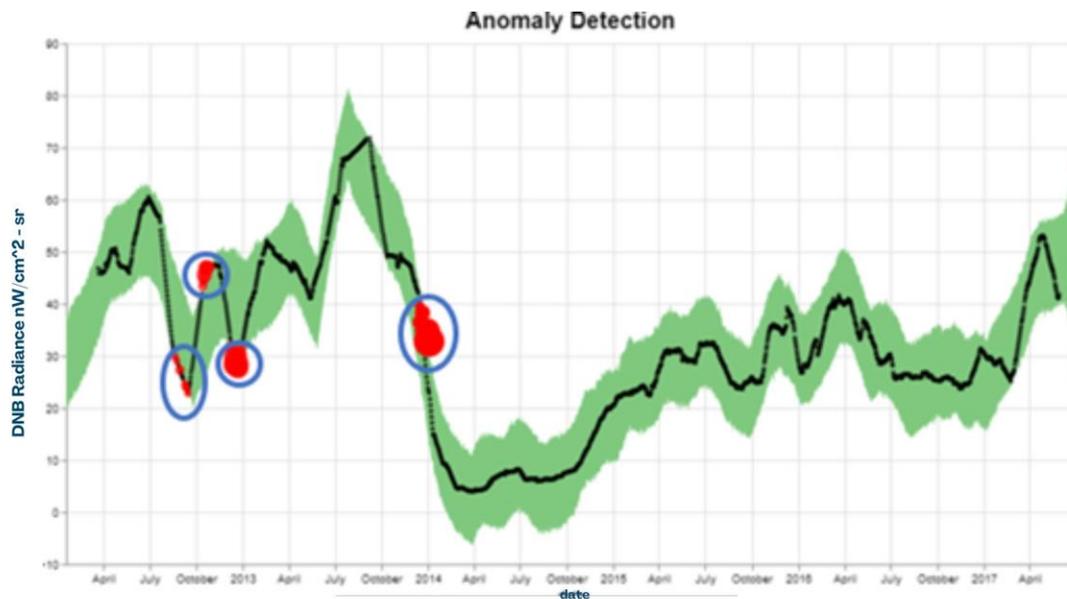
Figure 5. Time series cross-event correlation of anomalous values for the chosen pixel in Tacloban. Anomalies (red) were detected from fbprophet, wherein actual Nighlight values fall beyond the acceptable interval width (green). Four dates were seen as anomalies: August 2012, November 2012, December 2012, and November 2013.

From the image, several cases of anomalies were detected in the area. The first considered anomaly is the small red circles around the August 2012 timeline. We can see here that although the anomalous values are not that high, they were recorded as anomalies to that particular time, which coincides with the earthquake disaster mentioned earlier. The second anomaly was a spike, and specific values of nightlight radiance were higher than normal. Although this is not associated with any disasters, one may link this to the recovery phase of the said area. From the time that the city experienced a power outage up until this anomaly, the value is expected to jump high back up. Another possibility is the presence of false positives due to the chosen parameters of the model. The subsequent dip in values was recorded around December 2012. This again coincides with the second disaster mentioned earlier, the Typhoon Wukong. Lastly, the last area showing the red anomalies was dated around November to December 2013. This dip in nightlight values can be associated with the Super Typhoon Haiyan. The anomalies coincide with the dates of the disasters, as seen in the chart. As seen in the pixel-wise time series anomaly and event correlation, the anomalies detected were associated with the disasters that happened in the area. The pixel-wise model effectively provides insight into the time series relation with the disaster cases.

Another way of visualizing the effects of Super Typhoon Haiyan is to generate anomaly maps for each day. The anomalous values for all the pixels in a single day were recorded, and these points were mapped using QGIS. Anomaly maps for 2013 and 2014 were generated to visualize the locations of the anomaly points properly. Three of the several anomaly maps that were generated to visualize the effects of the Super Typhoon can be seen in Figure 6.
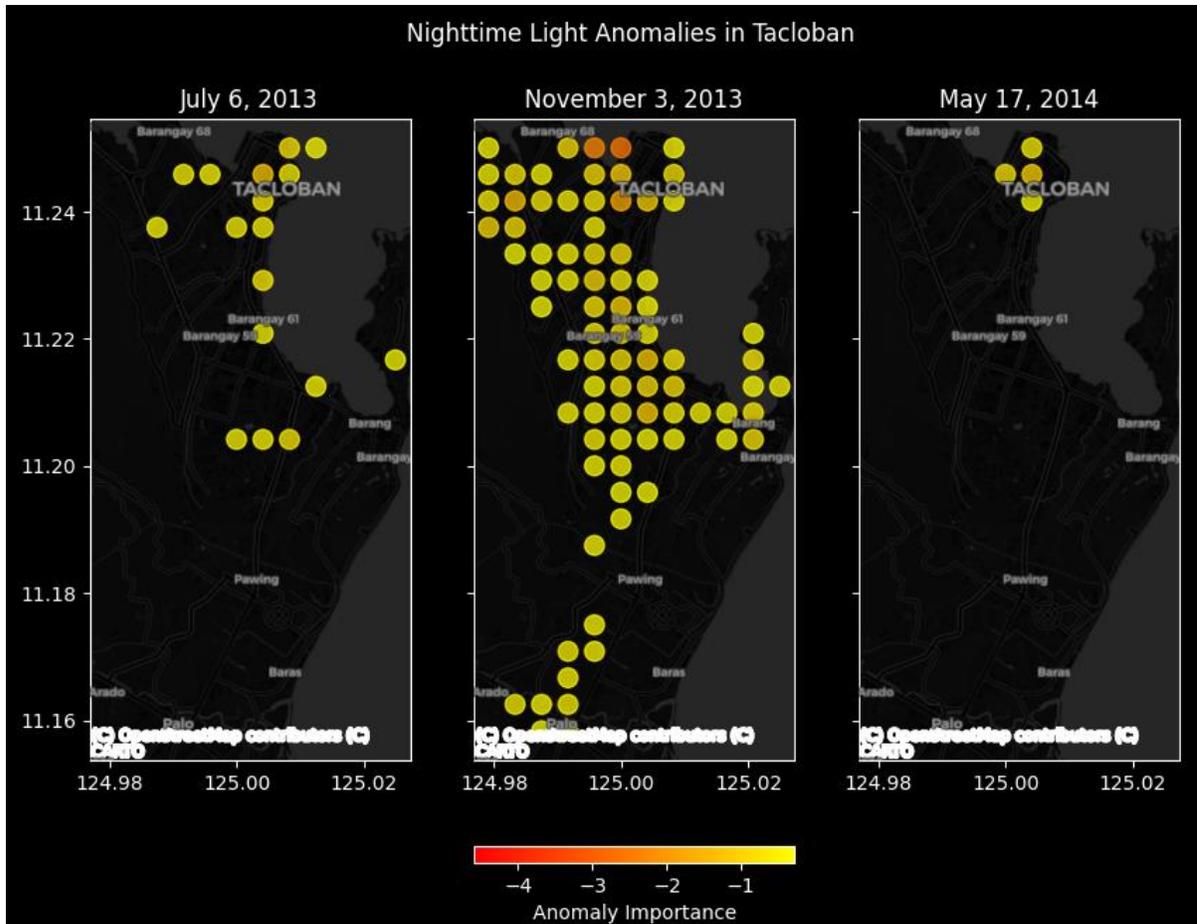
Figure 6. Anomaly maps for Tacloban for chosen dates: (a) July 6, 2013 (left); (b) November 3, 2013 (middle); and (c) May 17, 2014 (right). Circles show the location of nightlight anomaly, with its color intensity varying according to the calculated anomaly importance.

From the figure, the spatial extent of the typhoon can be observed by viewing the anomalous values of nighttime radiances. Significantly few anomalous values were observed in Figure 6a, which covered a date before the typhoon, and most of the area is not considered an anomaly. Possibilities such as these may be false positives, unaccounted power outage events, or simple nightlight deviation from the baseline pattern. Meanwhile, we can see in Figure 6b that most of the points within the city are anomalous, which means that during the actual day of the disaster, most of the city is affected by the typhoon. The extent of the effect of the typhoon can also be visualized. Lastly, Figure 6c shows the anomaly map for May 17, 2014, six months after the event happened. As evident from the map, the area started to clear, and no anomaly points were plotted. This is regarded as a recovery sign, in which case the city has gained access to electricity again, returning the nightlight's value to normal. The recovery period for Tacloban started in early February 2014, three months after the disaster. Full recovery happened after six months, in which no more anomaly points were visible.

From this case, we can see the value of the anomaly maps generated through the baseline modelling and anomaly detection algorithms. The algorithms have successfully detected the disasters in the area and determined the extent and recovery period of the said city.

3.2 Typhoon Vamco (2020) in Tuguegarao, Cagayan

Since the dates of Typhoon Vamco were beyond the available dates for VNP46A2, the dataset used for this case is VNP46A1. Several anomaly maps were generated, which covered dates before, during, and after Typhoon Vamco hit Tuguegarao City. Four of these are shown in Figure 7.
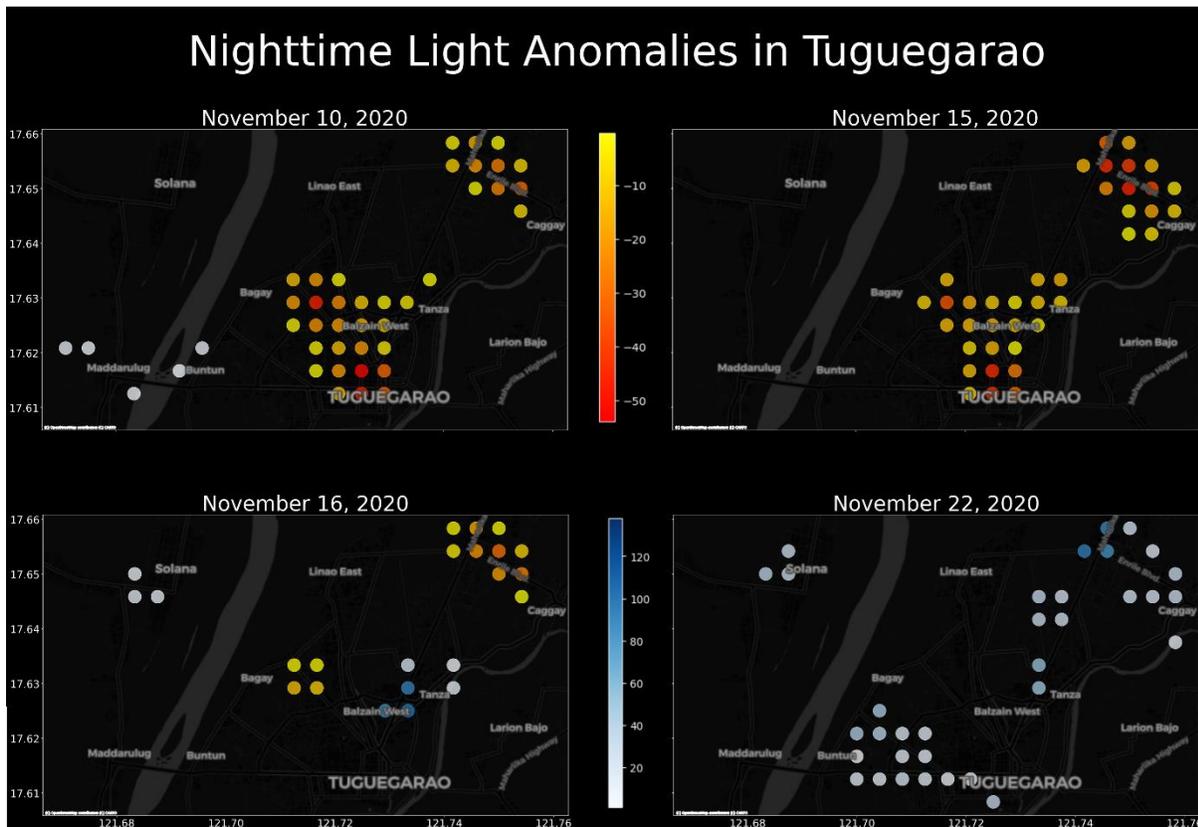

Figure 7. Anomaly maps of Tuguegarao, Cagayan during Typhoon Vamco last (a) November 10, 2020 (top left), (b) November 15, 2020 (top right), (c) November 16, 2020 (bottom left), and (d) November 22, 2020 (bottom right)

The area was heavily flooded even after the typhoon hit the country. The city was flooded for at least four days since the nearby river overflowed (Visaya, 2020). From the anomaly maps in Figure 7a and Figure 7b, most of the areas within the city were determined to be anomalous. The areas that were marked as red points were anomalous. From November 10 to November 15, the same regions were detected as anomalous areas. However, the anomalous points began to disappear for the succeeding days from November 16 to 22 as the area recovered, which is observable from Figure 7c and Figure 7d.

The blue points in Figure 7 correspond to areas with higher than baseline nightlight radiance values. In contrast, the red points represent areas lower than the interval width values calculated in the anomaly mapping. From November 16 to November 22, the number of red points began to decrease, while blue points were observed to replace the red points. As observed in Figure 7d, there were no more red points, meaning that the whole area had fully recovered and gained electricity by this time.

3.3. Extent of Disaster and Recovery

The two typhoon cases exhibited the model's potential in identifying the total extent of the effect of both typhoons. The anomaly mapping also provided a visualization of the recovery rate of the area during the disaster. For the case of Tuguegarao, Cagayan, the Typhoon Vamco onset was November 10, 2020. Negative nightlights were observed from November 10 to November 15, while positive ones were seen from November 16 to November 22. This means that the recovery rate for the area was two weeks after the typhoon hit. Meanwhile, the onset of Haiyan was on the 1st of November. The anomaly maps of Tacloban showed negative nightlights from November 2013 to February 2014. Recovery of the area was gradually observed from late February to April 2014. This means that the recovery period for the super typhoon Haiyan was much longer.

3.4. Comparison of VNP46A1 and VNP46A2

Both cases show the appropriateness of utilizing historical nightlight radiance datasets for identifying disaster-related anomalies. The daily temporal resolution of both datasets gives merit to its use for historical mapping and baselining. However, as mentioned earlier, the utilization of VNP46A1 provided less accurate information on the radiance values due to uncorrected moonlight. As seen in Figures 6-7, the range of anomaly importance values was higher in VNP46A1 than in VNP46A2. This may have been due to the uncorrected nightlight radiance values in VNP46A1, which led to a higher range of radiance values for the said dataset.

IV. CONCLUSION AND RECOMMENDATIONS

From the different tests and cases presented in this study, it has been shown that the baseline modelling and anomaly detection algorithm developed proved to help correlate relevant disasters to anomalous values in big datasets. The research successfully generated a model specifically for the VNP46A1 and VNP46A2 historical datasets that show promise in identifying anomalies as disasters in the historical dataset. Regarding the appropriateness of data, the VNP46A1 and VNP46A2 daily datasets provide ample historical data to provide a robust baseline model for nightlight radiance values.

The baseline model generated shows promise as it could model temporal patterns extracted from historical datasets. The baseline model has proven helpful in providing a quick assessment of the effect of certain events, like in the case of Typhoon Vamco and Haiyan, as demonstrated. The short evaluation using the baseline and the actual values can estimate the extent of the anomalous areas, providing helpful insights in scenarios such as disaster response.

Aside from these, anomaly mapping offers an excellent understanding of the timeline of the effects of certain events and the extent and timeline of the recovery of damaged areas. This was demonstrated in the two cases of disaster anomaly mapping.

Anomaly detection and baseline modelling were applied to several cases of disaster. The anomaly detection was successfully demonstrated to identify such disasters within the pixel-based time series, as shown by Typhoon Vamco and Haiyan cases. Refining the proposed anomaly detection is recommended to lessen instances of false positives, as demonstrated in some cases. Other variations to the algorithm can be applied, such as a mixture of spatiotemporal cases, wherein the values of nearby pixels could affect the values in the succeeding time series. Another suggestion would be to utilize other datasets that are similar in nature for VIIRS SNPP DNB. Finally, researchers may incorporate ancillary datasets, such as each pixel's land cover, to provide a customized baseline modelling and anomaly detection approach.

References:
[1]   Sanderson D. 2000. Cities, disasters, and livelihoods. Environment and Urbanization. 12(2):93-102.
[2]   McMichael A. 2000. The urban environment and health in a world of increasing globalization: issues for developing countries. Bulletin of the World Health Organization. 78(9):1117-1126.
[3]   McGranahan G, Songsore J, Kjellen M. 1996. Sustainability, the environment and urbanization: Sustainability, poverty and urban environmental transitions [e-book]. 1st ed. Routledge. http://dx.doi.org/10.4324/9781315800462-8
[4]   Chang S, Shizuoka M. 2004. Measuring improvements in the disaster resilience of communities. Earthquake Spectra. 20(3):739-755.
[5]   Joyce K, Wright K, Samsonova S, Ambrosia V. 2009. Remote sensing and the disaster management cycle [e-book]. Intech. doi: 10.5772/8341
[6]   Loor J, Fdez-Arroyabe P. 2019. Big data for remote sensing: Visualization, analysis and interpretation: Digital Earth and Smart Earth [e-book]. Springer Cham. https://doi.org/10.1007/978-3-319-89923-7
[7]   Madhukar M, Pooja. 2019. Big data for remote sensing: Visualization, analysis and interpretation: Earth Science [Big] Data Analytics [e-book]. Springer. https://doi.org/10.1007/978-3-319-89923-7_4
[8]   NASA. 2020. Retrieved from https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/5000/VNP46A1 on 11 Jan 2023.
[9]   Li T, Zhu Z, Wang Z, Roman M, Kalb V, Zhao Y. 2022. Continuous monitoring of nighttime light changes based on daily NASA's Black Marble product suite. Remote Sensing of Environment. 282:113269.
[10]   Levin N, Kyba C, Zhang Q, de Miguel A, Roman M, Li X, Portnov B, Molthan A, Jechow A, Miller S, Wang Z, Shrestha R, Elvidge C. 2020. Remote sensing of night lights: A review and an outlook for the future. remote sensing of environment. 237:111443.
[11]   Wang Z, Shrestha R, Roman M. 2020. NASA's Black marble nighttime lights product suite algorithm theoretical basis document (ATBD). Greenbelt, Maryland: NASA Goddard Space Flight Center.
[12]   Ahmed M, Mahmood A, Hu J. 2016. A survey of network anomaly detection techniques. Journal of Networks and Computer Applications. 60(C):19-31.
[13]   Agrawal S, Agrawal J. 2015. Survey on anomaly detection using data mining techniques. Procedia Computer Science. 60:708-713.
[14]   Thudumu S, Branch P, Jin J, Singh J. 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. Journal of Big Data. 7(42).
[15]   Zhou ZG, Tang P. 2016. Continuous anomaly detection in satellite image time series based on Z-scores of season-trend model residuals. 2016 IEEE International Geoscience and Remote Sensing Symposium

(IGARSS); Beijing, China. IEEE. p. 3410-3413. doi: 10.1109/IGARSS.2016.7729881.

[16] Verbesselt J, Zeiles A, Herold M. 2012. Near real-time disturbance detection using satellite image time series. Remote Sensing of Environment. 123:98-108.

[17] Kennedy R, Yang Z, Cohen W. 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. Remote Sensing of Environment. 114(12):2897-2910.

[18] Yin H, Pflugmacher D, Kennedy R, Sulla-Menashe D, Hostert P. 2014. Mapping annual land use and land cover changes using MODIS Time Series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7(8):3421-3427.

[19] Anees A, Aryal J. 2014. Near-real time detection of beetle infestation in pine forests using MODIS data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7(9):3713-3723.

[20] Chandola V, Vatsavai R. 2011. A Gaussian process based online change detection algorithm for monitoring periodic time series. 2011 SIAM International Conference on Data Mining; Mesa, Arizona. Society for Industrial and Applied Mathematics. p. 95-106. doi: 10.1137/1.9781611972818.9

[21] Grobler T, Ackermann E, van Zyl A, Olivier J, Kleynhans W, Salmon B. 2012. Using Page's cumulative sum test on MODIS time series to detect land-cover changes. IEEE Geoscience and Remote Sensing Letters. 10(2):332-336.

[22] Medium. 2020. Retrieved from https://medium.com/dataseries/anomaly-detection-in-satellite-imagery-e81fcf08abb8 on 11 Jan 2023.

[23] BBC News. 2012. Retrieved from https://www.bbc.com/news/world-asia-19435229 on 11 Jan 2023

[24] AFP. 2012. Retrieved from https://www.livemint.com/Politics/zo2LX68n9N0dI790Nt88HI/Tropical-storm-Wukong-threatens-Philippines.html on 11 Jan 2023

[25] The Atlantic. 2013. Retrieved from The Atlantic: https://www.theatlantic.com/photo/2013/11/super-typhoon-haiyan-devastates-the-philippines/100625/ on 11 Jan 2023

[26] CNN Philippines. 2020. Retrieved from https://www.cnnphilippines.com/news/2020/11/13/Ulysses-vs-Ondoy-comparison-NDRRMC.html on 11 Jan 2023

[27] Chang J, Hansen M, Pittman K, Carroll M, DiMiceli C. 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. Agron J. 1654.

[28] Chen B, Huang B, Chen L, Xu B. 2017. Spatially and temporally weighted regression: A novel method to produce continuous cloud-free landsat imagery. IEEE Transactions on Geoscience and Remote Sensing. 55(1):27-37.

[29] Shumway R, Stoffer D. 2017. Time series analysis and its applications with R examples [e-book]. 4th ed. Springer Cham. https://doi.org/10.1007/978-3-319-52452-8

[30] Inquirer. 2020. Retrieved from https://newsinfo.inquirer.net/1361922/floods-show-cagayan-environmental-risks on 11 Jan 2023.