9-29-2021

UNSUPERVISED LEARNING FOR ANOMALY DETECTION IN REMOTE SENSING IMAGERY

Husam A. Alfergani
Rowan University

UNSUPERVISED LEARNING FOR ANOMALY DETECTION IN REMOTE SENSING IMAGERY

by

Husam Alfergani

A Dissertation

Submitted to the
Department of Electrical and Computer Engineering
College of Engineering
In partial fulfillment of the requirement
For the degree of
Doctor of Philosophy
at
Rowan University
July 16, 2021

Dissertation Chair: Nazari Rouzbeh, Ph.D.

Committee Members:
Nidhal C. Bouaynaya, Ph.D.
Francis Haas, Ph.D.
Jess Everett, Ph.D.

Dedications

I dedicate this work to my parents. I made it this far following their wisdom they left behind, especially with their prayers of success for me throughout their lives.

Special dedication to my family, especially to my wife Khadija and kids Noha and Taha. Their love and support transcended the distance between us and they were with me every step of the way.

Finally, I dedicate this work to my country. I left on government scholarship awarded to me by The State of Libya to further expand my knowledge. Their continuous support throughout the years couldn't have been easy especially during the strenuous times of war and instability. May peace return to my country soon.

Acknowledgments

I would like to thank my advisors Dr. Nidhal Bouayanya for her continuous guidance and support and to Dr. Nazari Rouzbeh introducing me to the exciting field of remote sensing. I would also like to extend my deepest gratitude to the members of my committee for their fair and constructive feedback along the years.

I would also like to acknowledge the CBIE for the scholarship and funding since the start of this long journey despite the hardships that Libya is going through.

Most of all, I would like to extend a special thanks to my wife Khadija Ben Musa for her patience, fortitude and uplifting spirit that made this journey pleasant albeit the stress. Her support is the pillar of my success.

Abstract

Husam Alfergani
UNSUPERVISED LEARNING FOR ANOMALY DETECTION IN REMOTE
SENSING IMAGERY
2020-2021
Nazari Rouzbeh, Ph.D.
Doctor of Philosophy

Landfill fires is a potential hazard of waste mismanagement, and could occur both on and below the surface of active and closed sites. Timely identification of temperature anomalies is critical in monitoring and detecting landfill fires, to issue warnings that can help extinguish fires at early stages. The overarching objective of this research is to demonstrate the applicability and advantages of remote sensing data, coupled with machine learning techniques, to identify landfill thermal states that can lead to fire, in the absence of onsite observations. This dissertation proposed unsupervised learning techniques, notably variational auto-encoders (VAEs), to identify temperature anomalies from aerial landfill imagery. Twenty years of Landsat satellite observations at a number of landfills were examined for hotspots that may be associated with or leading to subsurface fires. The main contribution of this dissertation is to detect temperature anomalies in landfills using the state-of-the-art unsupervised deep learning technique of VAE based on both model reconstruction error and encoder module feature extraction. Additionally, a simple framework for assessing the health state of the landfill at any given time was established by using the clustering findings to generate a past behavior for each location in the landfill and eventually assigning it to one of four risk categories (No Risk, Low Risk, Moderate Risk and High Risk). This framework can function as a monitoring system, inferring information such as past landfill temperature profiles, predicting possible heat elevation or smoldering events as new observations are added, and identifying the percentage of each of the four risk categories and how they increase or decrease over the lifetime of the landfill.

Table of Contents

Table of Contents (Continued)

Table of Contents (Continued)

List of Figures

List of Tables

Chapter 1

Introduction

Remote sensing imagery acquired from satellites can be converted to land surface temperature (LST). The calculated LST can be analyzed to show the temperature variation within landfills. To validate the results obtained from studying heat elevation data using LST observations, it is essential to have ground truth data measured at the landfill. Unfortunately, not all landfill operators keep or publish heat elevation data and many landfills are not equipped with a landfill gas extraction system to control subsurface temperatures generated from the chemical reactions within. Hence, the calculated LST's are not supported by ground truth data that can be used to validate fire events or anomaly temperature areas within the landfill that should be controlled. To address the problem of the absence of onsite observations, one of the main goals of this study is the demonstration of the applicability and advantages of remote sensing data coupled with machine learning techniques necessary to identify landfill thermal states that can lead to fire events. On one hand, remote sensing can be used to locate hotspots by monitoring the thermal signature of these landfills. On the other hand, the machine learning algorithms will address the problem of the missing ground truth data "labeled data" by applying unsupervised machine learning methods to detect the thermal states of the landfills and to detect anomalies.

1.1 Unsupervised Machine Learning

Unsupervised machine learning has been defined as using "machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention." [1]. Therefore, unsupervised learning algorithms are self-learning without the need for any ground truth data (labeled data) and they will be able to find the relations in the given data. Unsupervised learning

is built on the idea of passing large volumes of unstructured data to algorithms or neural networks and enabling them to learn and infer from it [2].

1.2 Clustering

Clustering is an unsupervised machine learning technique that separates data into a predefined number of clusters, with observations of similar features are clustered together in one cluster. The most often used clustering approach is centroid clustering, also known as partitioning clustering, in which data points within the same cluster have the shortest distance [3]. Hard or exclusive clustering, is when each data point belongs to one and only one cluster. To achieve this for a given dataset, (1) data within the same cluster should have minimum distance, and (2) data of different clusters should have maximum distance [4].

1.3 K-Means Algorithm

K-means is one of the most commonly used unsupervised clustering algorithms, in which a given observation is allocated to a preset number of clusters k, in such a way that the clustered observations have maximum variance between clusters and minimum variation within the same cluster [2]. K-means is a partitioning-based clustering algorithm that organizes observations into k clusters based on distance measures [5]. It is an iterative clustering algorithm that attempts to find the local maxima by minimizing the objective function J in each iteration by assigning new clusters centroids [3, 6].

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|\mathbf{x}_i - \mu_i\|^2 \tag{1}$$

where $\|\mathbf{x}_i - \mu_i\|^2$, is the Euclidean distance (square norm) between the ith observation and the cluster center μ_i and C_i is the number of observation assigned to cluster i. Therefore for a given number of observations $x_1, ..., x_i$ for $x_i \in R^n$, the goal is to minimize the objective function, i.e. to minimize the Euclidean distance between the observation and its cluster

2

center C_i.

$$\underset{C_k}{\arg\min} \sum_{i=1}^{k} \sum_{x \in C_i} \|\mathbf{x}_i - \mu_i\|^2 \tag{2}$$

Where C_k is the kth cluster. If the distance used is Euclidean distance, then the distance from the observation to its centriod represents the variance. Equation 2 will calculate the distance of x_i to all clusters and assign the observation to the cluster with the lowest distance.

Equation 3 is used to update the centroid's positions after each iteration.

$$\mu_i = \frac{1}{C_i} \sum_{x \in C_i} \mathbf{x}_i \tag{3}$$

1.4 Anomaly Detection

Anomalies or outliers "are patterns in data that do not conform to a well-defined notion of normal behavior" (Chandola et al., 2009). Hence, the need for a process of identifying abnormal observation occurrences in unlabeled datasets that deviate from normal behavior. Unsupervised anomaly detection is frequently used since it does not require labeled data, which is rarely available. Labeling datasets is an expensive and time-consuming procedure. Unsupervised anomaly detection, on the other hand, is based on two fundamental assumptions:

- Anomalies occur seldom in comparison to regular cases in any dataset [7].

- They have considerably different characteristics than typical cases.

Therefore, using clustering as unsupervised anomaly detection will lead to another assumption that is normal data instances belong to large and dense clusters, while anomalies belong to small clusters [7, 8]. The output for an anomaly detection algorithm is in the form of scores to identify if it is a normal or an anomaly cluster.

1.5 Problem Statement and Background

Currently, there is no reliable and cost-effective method available in the United States (U.S.) for detecting and monitoring subsurface smoldering events (SSEs) and related thermal imbalances at U.S. landfills [9]. Such a method is needed as a timely warning tool for the identification of the location and spatiotemporal extent of subsurface "hotspots," while also aiding in the prevention or minimization of costly subsurface fires and thermal damage to liners and gas/leachate handling systems. The space borne remote sensing of landfill surface temperatures by thermal infrared sensing offers a promising approach. The interpretation of the publicly available Landsat data archive enables the monitoring of large areas, such as landfills. The nondestructive, noninvasive methods described in this work allow for the observation of multiple locations quickly and at low-to-no cost and the assembly of a satellite image archive that indicates changes in the thermal state of landfill surfaces over time. Further algorithmic interpretations of these thermal–areal time series can be used to isolate persistent hotspot signatures by filtering externally forced thermal variations (e.g., from seasonal thermal trends) and short-term thermal excursions [10].

Despite all the advantages of remote sensing data mentioned above, it is still lacking day by day ground truth data necessary to validate it. For instance, not all active landfill operators keep or publish heat elevation data, not to mention the closed, neglected and illegal waste sites where there are no data of any kind available. Furthermore, some landfills closed for years and still have some subsurface activity [11] .

Landfill subsurface heat is a normal and constantly active during the lifetime of a landfill. Subsurface smoldering events (SSE) that lead to surface fire do not happen momentarily, they have a long history that can extend to months and even years of continuous heat elevation. Therefore, a few remote sensing observations cannot tell us about the initi-

4

ations of these events, especially with the lack of ground truth data. Rather there must be a monitoring system that can shed a light on past events, estimate the initiation of current heat elevation events and be able to predict the future events.

1.6 Research Objectives and Contributions

In this work, we introduce methodologies for the remote satellite monitoring of the location and movement of subsurface thermal events within landfills, such as smolders and fires. As a case study, these methods were applied to the Bridgeton Sanitary Landfill in Bridgeton, Missouri, U.S., and several landfills in South New Jersey, U.S. Abnormal subsurface thermal activity has been ongoing at Bridgeton landfill since 2010 [12, 13, 14].

Considering that an anomaly is an unusual event that needs to be identified and monitored for the lifetime of landfills, and not as an outlier to be removed, the lack of labeled data from landfill operators imposes the use of unsupervised clustering methods for anomaly detection. Under these circumstances, we proposed the use of unsupervised deep learning Variationa Auto-Encoder (VAE) to extract low-dimensional salient features of the image from the encoder module of the VAE and feed them to a K-means clustering algorithm (VAE K-means) . VAE K-means is used to either cluster the thermal status of the landfill with K = 4 or to detect anomaly areas with K = 2.

With K = 4, the landfill is segmented into 4 areas labeled as no risk area, lower risk area, moderate risk area and higher risk area equivalent to clusters (1, 2, 3, and 4). The four areas are traced to 20 years back to build the temperature profile of the landfill. With K = 2, the landfill is segmented in two clusters, one cluster shows the anomaly area that exhibits the pixels with the highest temperature in the landfill, while the other cluster will show the remaining pixels of the landfill regardless of their thermal state

5

Another deep learning model based on VAE was proposed, where the VAE was trained to learn the distribution of normal data (without anomaly). When a new data with anomaly is fed to the model, the anomaly areas can be identified and localized using the reconstruction error. The results of this model can be compared to the previous model with $K = 2$.

The main contributions of this dissertation are to:

1. Leverage state-of-the-art unsupervised deep learning method of VAE to detect temperature anomalies in landfills based on both the reconstruction error of the model and feature extraction of the encoder module.

2. Quantify the health status of the landfill at any given time using the clustering results to build a historical behavior for each region in the landfill, and ultimately to label it to one of the four categories (No Risk, Lower Risk, Moderate Risk and Higher Risk). This framework is in the form of tables and graphs that constitute a monitoring system, where the following information can be inferred:

   - The past landfill temperatures profile.

   - Predict the possible heat elevation that may lead to a smoldering events as we keep adding new observations.

   - Identify the percentage of each of the four categories (No Risk, Lower Risk, Moderate Risk and Higher Risk) and how they increase/decrease along the lifetime of the landfill.

1.7 Organization of the Dissertation

In chapter 2, we reviewed the primary cause of landfill heat elevation as well as the applicability of remote sensing data in detecting temperature changes. In addition, we discussed the statistical approach in anomaly detection and change detection. We reviewed

6

the state-of-the-art unsupervised data-driven models for anomaly detection utilizing deep learning models. Finally, we highlighted the important factors that must be considered when detecting anomalies in landfills.

We dedicated chapter 3 to Satellite Data Acquisition and Processing. First acquiring satellite imagery for the areas of interest, then we showed the steps of calculating Land Surface Temperature (LST). Second, each LST observation is reduced to a suitable size for efficient processing time and memory use. During the processing of LST observation we generated multiple data formats such as: images, temperature data in TIF files, and videos that are used throughout this work. Furthermore, the resulting images are used to form two datasets to train variational autoencoders (VAE) in chapter 4.

In chapter 4, we introduced a new approach to the problem of change detection in LST remote sensing observations based on statistical sequential analysis theory. The reduced dimensionality aims at increasing the computational efficiency taking into account the large size of remote sensing data. The statistical approach is based on detecting the change in the mean parameter of the generative distribution of the stochastic data.

In chapter 5, we explored landfill heat elevation from different perspectives. First we analyzed temporal trends of the landfill to look for any signs of heat elevation. Then we used unsupervised K-means clustering and Heat Index (HI) to study the spatial temperature trends, and to cluster the landfill into four thermal states (no, lower, moderate, and higher risk areas) for each observation. Finally, we introduced the idea of accumulated indexes and Frequency of Maxima (FM) as spatio-temporal analysis to shed light on the past history of heat elevation in landfills.

In chapter 6, we proposed using state-of-the-art unsupervised deep learning VAE to

7

detect and localize temperature anomaly in landfills based on both the reconstruction error of the model and feature extraction of the encoder module.

In chapter 7, we evaluated the clustering results using quantitative methods, the Elbow and the Silhouette as one of the most common internal cluster validation methods.

In chapter 8, we proposed a simplified framework to quantify the health state of the landfill based on spatio-temporal analysis. The quantification of the health state of the landfill can evaluate the current state of the landfill, past events and predict where the next heat elevation or possible fire may occur.

Chapter 9, provides a brief summary and conclusions of this dissertation.

Chapter 2

Literature Review

The US Federal Emergency Management Agency (FEMA) reports that approximately \$8.4 million dollars of yearly damage to property is caused by landfill fires [15, 16]. There is a need for a timely warning tool that can identify location and spatiotemporal extent of subsurface "hotspots", perhaps aiding in prevention or minimization of costly subsurface fires or thermal damage to liners and gas/leakage handling systems. To date, a scalable, cost-effective, and reliable method to detect and monitor subsurface smoldering events and related thermal imbalances at landfills remains broadly lacking [9, 17, 14].

2.1 Reasons Behind Landfill Heat Elevation

Biological degradation of organic and chemical waste is one of the main reasons for temperature elevation in landfills. The degradation process passes through aerobic (characterized by high percentage of oxygen) and anaerobic phases (characterized by the depletion of oxygen and the existence of other gases such as methane and CO_2) once a given mass of waste is deposited in a landfill [15, 18, 19]. The byproduct of this process is heat, leachate, and gas [20, 21, 22, 23] where rates of heat generation are higher during aerobic phases [24, 25, 26, 27].

A case study found that landfill gas under normal operating conditions is at atmospheric pressure and at a temperature of 40°C [28]. If subsurface temperature is elevated to 100°C, the pressure would rise by approximately 20 kPa, and allow hot gasses to accumulate by convection under the surface of the landfill. Under normal operating conditions, the landfill temperature remains close to the air temperature at shallow depths and near the edges of the landfill and reaches maximum values relative to the air and ground temperatures near the areal center and at intermediate depths.

9

2.2 Application of Remote Sensing in Temperature Change Detection

Previous studies indicated that a baseline "healthy" landfill thermal state can be observed by satellite-based remote sensing [14]. Continuous and long-lasting subsurface heating activities result in higher surface temperatures by the transfer of heat from the interior to the landfill surface [29, 10, 30]. Data from landfills experiencing SSEs, subsurface oxidation events, or elevated temperatures suggest that temperatures inside landfills can reach 150 °C [28]. Other studies indicated that satellite-based remote sensing applications can identify and map landfill sites based on differences between surface temperatures and their surroundings. Remote sensing was used to map and monitor Al-Qurain landfill in Kuwait, where temperature differences up to 4 °C were observed from the surrounding desert area [31]. At the Trail Road landfill site near Ottawa, Canada, Kwarting and Al-Enezi observed up to 9 °C and 14 °C temperature difference between surrounding areas and air temperature from 1985 to 2009 [32]. However, none of these studies focused on the detection and monitoring of persistent hotspots (anomaly temperatures) as an indicator of landfill health disturbance.

2.3 Statistical Approach in Temperature Change Detection

Recently, we proposed a statistical on-line change detection algorithm [33]. We formulated the problem of spatio-temporal Land Surface Temperature (LST) detection as a statistical sequential change detection problem. LST images are modeled as stochastic processes, with temperature changes reflected as changes in the mean parameter of the process. A dimensionality reduction using Direct Cosine Transformation (DCT) followed by Principal Component Analysis (PCA) was used to increase the computational efficiency given the large size of remote sensing datasets. The results show that this approach can detect gradual and abrupt changes in the landfill and as a special case, it was able to detect anomaly changes compared to the mean taken over a large area of landfill. Statistical

10

sequential analysis was investigated for temporal video scene detection in [34] and spatial detection of macrocalcification in digital mammograms in [35].

Previous research using remote sensing of landfill surface temperatures using statistical approaches focused on detecting coalmine fires from satellite imagery. Deterministic techniques that rely on setting a detection threshold were used in [36, 37]. This approach depends on previous knowledge of the terrain. Dynamic thresholding techniques [38, 39] for subsurface coal fires detection use histogram analysis and varying-size moving window methods. They define a threshold as the first local minimum after the local maximum of the histogram within each window. A pixel is marked as "thermally anomalous" if it is detected 70% of the time considering all window sizes. This process is followed by the 8-neighbours clustering method to identify the coal fire maps. A Gaussian process-based online detection algorithm was used in [40] and reported 78% accuracy in detecting change in the normalized difference vegetation index.

2.4   Unsupervised Data Driven Approach in Anomaly Detection

Recent research using deep learning showed promise in anomaly detection in various fields. For instance, Generative Adversarial Networks (GAN) were used in detecting and localizing forgery in satellite images, which were watermarked by foreign objects of different sizes [41]. The study assumed no forged images were available for training. However, the dataset used was small, i.e., 130 satellite images, including 30 images for training and 100 images for validation. Fifty of the validation images were forged. A variational autoencoder (VAE) was used for unsupervised anomaly detection based on feature extraction [42]. The extracted features were fed to different traditional unsupervised anomaly detection methods: K-Nearest Neighbor (KNN), Local Outlier Factor (LOF), Cluster-Based Local Outlier Factor (CBLOF), and One-Class Support Vector Machine (OC-SVM). The VAE was trained with the MNIST dataset. The authors reported an Area Under the Curve

11

(AUC) of 0.973 and 0.971 with KNN and OC-SVM using 900 -digit "7"- as normal data and 100 random samples from the rest of the digits. Sabokrou et al. proposed an end-to-end unsupervised model for a one-class classifier [43]. The model comprises of two networks. The first is an autoencoder that learns to reconstruct the input with minimum error, and the second is a discriminator with a scalar output (0,1) that generates an anomaly score. Both networks are jointly trained to learn the distribution of anomaly-free data. The discriminator network is then used at test time to classify the anomaly input. The model was tested on the Caltech-256 dataset and achieved an AUC and F1 of 0.942 and 0.928, respectively.

2.5 Considerations When Detecting Anomalies in Landfills

There are a few considerations when detecting anomalies in landfills:

1. Anomaly is an unusual event that needs to be identified and monitored for the lifetime of landfills and not as an outlier to be removed.

2. The nature of heat elevation in landfills dictates that multiple anomaly areas can exist in the same observation; hence, the detection problem should consider every observation in small patches of reasonable size.

3. The lack of labeled data from landfill operators imposes the use of unsupervised classification methods for anomaly detection.

In this work, we propose using unsupervised deep learning VAE in two ways to detect temperature anomaly in landfills. In the first method, we train a VAE to learn the distribution of normal data (without anomaly). When a new data with anomaly is fed to the model, the anomaly areas can be identified and localized using the reconstruction error. In the second method, we use the encoder module of the VAE to extract low-dimensional salient features of the image and feed them to a K-means clustering algorithm (VAE K-means). VAE K-means is used to either cluster the thermal status of the landfill with K = 4 or detect anomaly areas with K = 2. Although, the methods use different techniques for anomaly

12

detection, VAE K-means of K = 2 can be compared to the first method based on reconstruction error as both methods only detect the anomalies without providing information on the surrounding area. The second method with VAE K-means of K=4 provides a clearer view of the thermal status of the anomaly area and surrounding in the landfill. The individual clustering results are tracked and accumulated over time to assess the health status of the landfill.

Satellite Data Acquisition and Processing

3.1   Data Acquisition

The U.S. Geological Survey (USGS) Earth Explorer tool provides the ability to query, search, and order satellite images, aerial photographs, and cartographic products from several sources. However, none of these representations contain information related to temperature. To determine land surface temperature (LST) distributions, Level 1 satellite images of the exact location of the landfills (GeoTIFF format) were downloaded from the USGS online archive (https://earthexplorer.usgs.gov) and were then processed as described below. Observations from Landsat satellites were used to detect the thermal state and to identify thermal anomalies at its surface for the following landfills:

- Bridgeton Sanitary Landfill in Bridgeton, MO, USA between 2000-2016.

- Nine landfills in South New Jersey from 2000-2019 as shown in Table 1.

All relevant Landsat data for the dates shown above were downloaded and images with only slight (10%) overall cloud contamination were retained. All retained images were then subjected to an image acceptance test, i.e., an algorithm designed to use the Quality Assessment band (now available with downloaded data for Landsat 5, 7 and 8) to accept only images (in this study, the landfill scene) that have no clouds, snow, water, or other land cover that may lead to misleading results. In addition, the images were visually checked to ensure that the landfill area was not obscured by clouds. No reliable data were available between December 2011 and March 2013 as the Landsat 5 archive ends in November 2011, Landsat 8 was launched in April 2013, and the Landsat 7 data for 2012 were found to be unusable for this analysis because of sensor problems. The missing 2012 data is unfortunate, but hotspots can still be tracked for over 93% of the period of interest.

14

Table 1

The Location and Profile of the Selected Landfills

Landfill	County	Latitude	Longitude	Status	No. Images
Carney's Point Township	Salem	39.7030	-75.4868	closed	112
Commercial Township	Salem	39.2981	-75.0422	closed	112
Egg Harbor Township	Atlantic	39.427	-74.5376	active	112
Millville City	Cumberland	39.3832	-75.0563	active	112
Vineland City	Cumberland	39.7030	-75.4868	closed	113
Woodbine Borough	Cape May	39.2377	-74.7858	active	112
South Harrison Township	Gloucester	39.71	-75.285	closed	112
Deerfield Township	Cumberland	39.452	-75.100	active	113
Mannington Township	Salem	39.589	-75.372	active	112

Images from 2000–2011 were obtained using the Landsat 5 TM. Starting from 2013, we acquired data from the Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments. The number and positions of the spectral bands in the Landsat sensors differed, but we used all sensors that provided observations in the visible, near-infrared (near-IR), and thermal infrared (TIR) bands. The spatial resolution of all the sensors in the visible and near-IR bands was 30 m, and that of the TIR band was 120 m on the TM sensor and 100 m on the TIRS sensor. However, the USGS provides observations in these bands resampled to 30-m resolution, which is the same as that of the visible and near-IR bands. All scenes were acquired at Level 1B with observations in all bands provided as 8 bits for the TM and 16 bits for the OLI and TIRS.

15

3.2 Data Processing to Calculate Land Surface Temperature (LST)

Calculatin LST from the satellite imagery was performed according to the procedure described in the Landsat handbook [44]. Digital numbers (DNs) in the optical bands were converted first to radiance and then to reflectance. The reflectance values were corrected for variable Sun–Earth distances and normalized to the overhead Sun by dividing the reflectance by the cosine of the solar zenith angle. Observations in the TIR band were converted first to radiance and then to brightness temperature T_b values. The calibration coefficients used to convert DN counts into physical values (reflectance and brightness temperature) were obtained from metafiles supplied by the USGS with the Landsat imagery Table 2.

Table 2

Brightness Temperature Constant Values for Use with Equation 4

Constant	K1	K2
Units	W/(sq. m^2 μm)	Kelvin
L5 TM	607.76	1260.56
L8 TIR	774.89	1321.08

To estimate the Land Surface Temperature (LST) from the observed IR brightness temperature T_b, we used the approach below [45]:

$$LST = \frac{T_b}{1 + \left(\frac{\lambda T_b}{d}\right)\ln(e)},\tag{4}$$

where T_b is the black body temperature; λ is the wavelength of the emitted radiance; d is

16

defined by $d = ch/k_B$, where the velocity of light ($c = 3 \times 10^8$ m/s) is multiplied by Planck's constant ($h = 6.26 \times 10^{-34}$ J.s) and divided by Boltzmann's constant ($k_B = 1.38 \times 10^{-23}$ J/K); and e is the land surface emissivity. The emissivity is calculated using Equation 5 [46]:

$$e = 0.004 + 0.986 P_v, \tag{5}$$

where P_V is the proportion of vegetation, which is sometimes referred to as the fractional vegetation cover and calculated using Equation 6:

$$P_V = \left[\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right]^2, \tag{6}$$

In this equation, NDVI is the normalized difference vegetation index. To calculate the NDVI of the surface, we used Equation 7 [47]:

$$NDVI = \left[\frac{NIR - RED}{NIR + RED} \right]. \tag{7}$$

where NIR represents the near-IR band reflectance and RED is the visible red band reflectance. $NDVI_{max}$ and $NDVI_{min}$ in Equation 6 are the maximum and minimum NDVI indices in the image, respectively, for which $NDVI_{max} = 0.5$ for vegetation and $NDVI_{min} = 0.2$ for soil can be used [48].

Atmospheric scattering and absorption may also affect the estimation of land surface emissivity from NDVI [49]. In this study, the effects of scattering and absorption on the NDVI—particularly the atmospheric absorption by carbon dioxide and water vapor—were not taken into consideration. This could result in an overall underestimation of the absolute LST by 1–3 K. However, it would not affect the contrasts and gradients of the observed surface temperature because (1) there is no reason to expect meaningful variations in the atmospheric composition over the area of a landfill and (2) such uncertainties

17

relative to the nominal absolute temperatures, which were subsequently considered, were not found to meaningfully impact hot spot detection.

3.3 Data Post-Processing

All downloaded satellite images (observation) for South New Jersey lay in a single tile ($path = 14$ and $row = 33$). Landsat 8 scene size is approximately $185\,km \times 180\,km$ and Landsat 5 is $170\,km \times 183\,km$. Using in-house MATLAB code, all images were read with 1-km buffer around a central point of a landfill (the Lat and Lon values given in Table 2) which covers approximately $4\,km^2$ for all bands, including the thermal bands, to allow studying and analyzing the landfill's surrounding area when needed. All observations are masked with an overall landfill shape file drawn on the border of the landfill. This approach considers only the readings within the area of interest and has two main advantages:

1. to eliminate the effect of any extreme or misleading reading due to water bodies, buildings, or any other land cover.

2. to reduce the LST image size from $185\,km \times 180\,km$ to the area of the landfill, which greatly decreases the processing time and reduce the memory requirements.

3.4 Data Formats

3.4.1 *Image Data*

The results obtained from post processing in the previous section were grouped and saved separately for every landfill in one of the following formats:

1. Temperature color coded (PNG/JEPG) images.

2. Raster[1] data (TIF) of real LST measured at the landfills.

---

[1] A raster as defined by ArcGIS- ESRI® is "a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as temperature. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps." [https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm].

18

All (TIF) files are Geo-referenced so that they can be overlaid on Google Earth to easily visualize the thermal stale of the landfill. These files were used in chapter 5 and chapter 7.

3.4.2   Video Data

All (PNG/JEPG) are compiled in video files (MPG4) as listed in Appendix A. For instance, LST observations of Bridgeton Missouri landfill between 2000 and 2016 were considered, and a total of 115 images compiled in (MPG4) video file.[2] This file is used in chapter 4 to simulate on-line temperate change detection in Bridgeton Missouri landfill.

3.4.3   Datasets for Anomaly Detection Using VAE

LST observations described in (PNG) format discribed in 3.4.1 were used to form datasets for both clustering and anomaly detection in landfills using variational autoencoders.

Dataset-1: consists of all observations of the nine landfills (800 RGB images), 600 for training and 200 divided between validation and testing. This dataset was used for anomaly detection using feature extraction and VAE K-means clustering. Dataset-1 is intended to extract the features of the colors representing the different temperatures in every observation.

Dataset-2: contains the original images with a small area of anomalies (red) or near anomaly (orange), and the rest are customized where the red color has been removed. This dataset consists of 110 images (90 for training and 20 for validation) and was used for anomaly detection and localization based on reconstruction error. This dataset is used to train the model to learn the distribution of normal data so it can detect any anomalies during testing. These datasets were used in chapter 6 as an input to the variational autoencoder.

---

[2][https://www.youtube.com/watch?v=YbOwkt0fH5k].

Chapter 4
Spatio-Temporal Statistical Sequential Analysis for Temperature Change Detection in Satellite Imagery

The analysis of remote sensing data enables us to detect changes and monitor land surface temperature (LST). However, analysis of times series data poses some challenges, including weather conditions, seasonality and noise, that limit the effectiveness of change detection algorithms. While existing algorithms perform relatively well for detecting abrupt transitions, reliable detection of gradual changes is more difficult. In this chapter, we formulate the problem of spatio-temporal LST change detection as a statistical sequential change detection problem. LST images are modeled as stochastic processes, with temperature changes reflected as changes in the parameters (i.e., mean) of the process. A generalized likelihood ratio test is used to detect these changes and estimate the exact time/space where they occur. To minimize processing time and memory requirements, we represent LST images by their reduced dimensionality using direct cosine transformation followed by principal component analysis. Statistical sequential analysis is used to provide a unified mathematical framework for the detection of both abrupt and gradual changes in LST observations of Bridgeton Missouri landfill over 17 years [33].

The statistical sequential analysis is considered for both spatial and temporal change detection of temperature from Landsat Thematic Mapper (TM) images. We transform a satellite image time series into a video, not only for visualization, but also for processing the satellite image series. TM observations are one of the most widely used datasets for environmental studies. LST images are first mapped to a lower dimensional space using Discrete Cosine Transform (DCT) followed by Principal Component Analysis (PCA). The image features, being the PCA eigenvectors, are considered as a realization of a stochastic

20

process and a change is detected if the mean parameter of the probability density function changes.

As mentioned in chapter 3.4, we compiled all observations into a video file.[1] The video file is intended to simulate online detection that is required for additive modeling and sequential statistical analysis. In the following section we will define these terms in the context of this chapter.

4.1 Basic Definitions

4.1.1 Change Detection Using Remote Sensing

Change detection is the process of measuring the change of the characteristics of a specified area among multiple time frames. Space or air remote sensing imagery such as satellites, aerial photography or Unmanned Aerial Vehicles (UAV's) are one of the main tools used for change detection; for the purpose of environmental monitoring of temperature changes and urban changes [50].

4.1.2 Online/Real-Time Change Detection

Online change detection algorithms operate in parallel with the process being monitored and evaluate each observation as it becomes available with the objective of detecting a change as soon as it occurs and just before the next observation becomes available [51, 52, 53].

4.1.3 Additive Modeling

Additive modeling implies that the detection algorithm sequentially reads the video frames of one scene, which represents the thermal status of the landfill. The sequential assumption in temporal change detection is based on the fact that we do not have the data

---

[1][https://www.youtube.com/watch?v=YbOwkt0fH5k].

21

available during processing, but we obtain one observation at a time as they become available (as video scenes are read) [33].

4.1.4 Sequential Statistical Change Detection

The LST change detection is modeled as an additive change in the mean (μ) parameter of the probability distribution function (PDF) characterizing the stochastic process. A hypothesis test to the the mean parameters before and after the change determines whether to (1) accept the change and stop reading more observation and wait for the next observation, or (2) reject it and continue sampling i.e., to read more observation until a change is detected. The change detection is accepted based on a comparing the test statistic to a given threshold [54, 55].

4.2 Dimensionality Reduction and Feature Vector Representation

To minimize the processing time and memory requirements, we represent the LST images by their reduced dimensionality. For spatial detection, we divide each image into macroblocks of size $M \times M$. The first step to dimensionality reduction is to compute the DCT as shown in Figure 1. A feature vector of DC coefficients, $\{X_k\}_{k=1}^N$, is formed by lexicographical ordering of the coefficients. We now consider a data matrix $P \in R^{N \times M}$, where M denotes the number of macroblocks in the spatial case and denotes the number of frames in the temporal case. Then, the eigenvalues and eigenvectors of the correlation matrix $C = P^T P$ are calculated. Thus, each DC vector X_k will be reduced to a feature vector Y_k obtained by projecting X_k onto a subspace of eigenvectors corresponding to highest eigenvalues. That is, every spatial macroblock is represented by this feature vector. Feature vectors $\{Y_k\}$ are passed sequentially to the algorithm until a change is detected. This process is depicted in Figure 1. In this application, we found that the largest eigenvalue accounts for more than 96% of the total eigenvalues of the correlation matrix. A detailed example is give in Appendix B to show the steps of dimensionality reduction as descried in

22

Algorithm 1

Figure 1

Dimensionality Reduction Using Discrete Cosine Transform Followed by Principle Component Analysis (PCA)

Total Reduction (Preprocessing) = (900 X 1200)/6864 ≈ 158 times

4.3 Statistical Sequential Analysis

Feature vectors $\{Y_k\}$ are assumed to form an independent and identical distribution (i.i.d.) sequence of r-dimensional random vectors $\{Y_k\}_{k>1}$ with Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ with pdf:

$$p_{\mu,\Sigma}(Y_k) = \frac{1}{\sqrt{(2\pi)^r|\Sigma|}} e^{(-\frac{1}{2})(Y_k-\mu)^T\Sigma^{-1}(Y_k-\mu)}. \tag{8}$$

Additive modeling implies that the detection algorithm sequentially reads the video frames of one scene, which represents the thermal status of the landfill in its reduced dimensionality form $\{Y_k\}$. The sequential assumption in temporal change detection is based on the fact that we do not have the data available during processing, but we obtain one observation at a time as they become available (as video scenes are read). Therefore, The LST change detection is modeled as an additive change in the mean vector parameter $\theta = \mu$ of the pdf characterizing the stochastic feature vector. If the parameter $\theta = \theta_0$ is the value before the change, then $\theta = \theta_1$ will be the value after the change. Because of the sequential assumption, we assume that we have no information about the parameter θ_1 after change.

23

Considering the case of the upper bound for θ_0 and lower bound for θ_1 are known, the change detection problem is equivalent to the following hypothesis:

$$H_0 = \{\theta : \|\theta - \theta_0\|_\Sigma^2 \leq a^2, k < t_0\}$$
$$H_1 = \{\theta : \|\theta - \theta_0\|_\Sigma^2 \geq b^2, k \geq t_0\}, \tag{9}$$

where $\|\theta - \theta_0\|_\Sigma^2 = (\theta - \theta_0)^T \Sigma^{-1}(\theta - \theta_0)$, k is the discrete time index, t_0 is the true change time and $a < b$, where a and b are the lower and upper bound for the change magnitude respectively. The solution to the detection problem in Equation 9 can be obtained by deriving the generalized likelihood ratio (GLR) test [55], where the unknown parameters are replaced by their maximum likelihood estimates. The generalized likelihood ratio for the sequence $\{Y_j, \ldots, Y_k\}$ is:

$$S_j^k = \ln \frac{\sup_{\|\theta - \theta_0\|_\Sigma \geq b} p_\theta(Y_j, \ldots Y_k)}{\sup_{\|\theta - \theta_0\|_\Sigma \leq a} p_\theta(Y_j, \ldots Y_k)}, \tag{10}$$

where p_θ is the parameterized probability density function. The sequential GLR algorithm is then given by:

$$t_a = \min\{k \geq 1 : g_k \geq h\},$$
$$g_k = \max_{1 \leq j \leq k} S_j^k, \tag{11}$$

where t_a is the alarm (detection) event, g_k is the test statistic, and h is a threshold. Given the i.i.d. Gaussian assumption, S_j^k can be written as:

$$S_j^k = \ln \frac{\sup_{\|\theta - \theta_0\|_\Sigma \geq b} e^{-\Sigma_{i=j}^k (Y_i - \theta)^T \Sigma^{-1}(Y_i - \theta)}}{\sup_{\|\theta - \theta_0\|_\Sigma \leq a} e^{-\Sigma_{i=j}^k (Y_i - \theta)^T \Sigma^{-1}(Y_i - \theta)}} \tag{12}$$

24

It can be shown that S_j^k can be rewritten as [55]:

$$S_j^k = \begin{cases} -(\chi_j^k - b)^2, & \chi_j^k < a \\ -(\chi_j^k - b)^2 + (\chi_j^k - a)^2, & a \leq \chi_j^k \leq b \\ +(\chi_j^k - a)^2, & \chi_j^k > b \end{cases} \tag{13}$$

where

$$\chi_j^k = \left[(\bar{Y}_j^k - \theta_0)^T \Sigma^{-1} (\bar{Y}_j^k - \theta_0) \right]^{\frac{1}{2}},$$
$$\bar{Y}_j^k = \frac{1}{k-j+1} \Sigma_{i=j}^k Y_i. \tag{14}$$

The data needed in Equation 14 are the feature vectors Y_i, the covariance Σ and the mean before the change θ_0. In the more realistic case where the parameter before the change θ_0 is assumed to be known but the parameter after the change is assumed to be completely unknown, the change detection problem statement is as follows:

$$H_0 = \{\theta : \theta = \theta_0, k < t_0\}$$
$$H_1 = \{\theta : \theta \neq \theta_0, k \geq t_0\}. \tag{15}$$

Hence, the case where nothing is known about θ_1 can be considered as the limit of the previous case when $a = b = 0$. Therefore, the GLR algorithm in Equation 11 becomes:

$$t_a = \min\{k \geq 1 : g_k \geq h\},$$
$$g_k = \max_{1 \leq j \leq k} \left\{ \frac{k-j+1}{2} (\chi_j^k)^2 \right\} \tag{16}$$

where χ_j^k is defined in Equation 14. In the above study, θ_0 is assumed to be known and can be estimated using the first M feature vectors of each LST video observation. The covariance Σ is estimated using the same M feature vectors. Algorithm 1, summarizes the the mains steps of LST change detection.

25

Algorithm 1: Sequential change detection algorithm

Input: Video file

1 **while** $\sim EOF$ **do**

2     $k := 1$ `// Reset for every new scene`

3     $change := false$

    `/* initialize phase`   `*/`

4   **while** $k \leq M$ **do**

5        $get\ X_k$ `// accumulate` X_k `until` $k := M$

6   **if** $k := M$ **then**

7           $Compute\ PCA, \Phi, \{Y_1, \ldots, Y_M\}$ `// Using Equation 27 in Appendix B`

8           $Estimate\ \theta_o, \Sigma$

                             `//` θ_o, Σ `before the change using` $\{Y_1, \ldots, Y_M\}$

9   **else**

10           $k := k+1$

    `/* Triggering the detection Algorithm`   `*/`

11   **while** $\sim change$ **do**

12        $get\ X_k$ `// New frame` X_k

13        $Y_k = \Phi^T X_k$ `// project` X_k `onto a subspace of eigenvectors Equation 28`

14        $Compute\ g_k$ `// test statistic` *Equation* `16`

15   **if** $g_k \leq h$ **then**

16           $k := k+1$

17   **else**

18           $change := true$ `// A change is detected`

19           $K := 1$ `// Restart Algorithm go to step 1 take new scene`

4.4 Simulation Results

To assess the performance of the proposed spatio-temporal statistical sequential analysis algorithm, we used 17 years of LST observations of Bridgeton Landfill, which correspond to 115 images.

Temporal Change Detection: To evaluate the temporal change detection, we replicated each image 14 times, i.e., 1 original + 14 replicas = 15, to create a video scene of 1725 frames (115x15). Note that each scene corresponds to 15 frames as shown in Figure 2a. Our aim is to test the change detection of every scene under noiseless and noisy conditions. The noisy conditions correspond to adding Gaussian noise to each replica with varying sig-

26

nal to noise ratio Figure 2b.

In the noiseless case, the temporal change detection algorithm is able to detect 100% of the changes. To assess the robustness of the algorithm, we added Gaussian noise to the replicated frames with variance $\sigma^2 = 0.01$. We used $M = 6$ to estimate the parameter $\theta_0 = \mu$ before the change. The detection algorithm sequentially finds the decision function g_k above the threshold h. We found experimentally that setting $h = 0.45$ best balances false positives and false negatives for this dataset.

Figure 2

Simulating Temporal Change Detection. Each Observation is Replicated 14 Times to Form a Homogeneous Scene; (a) No Noise; (b) Gaussian Noise
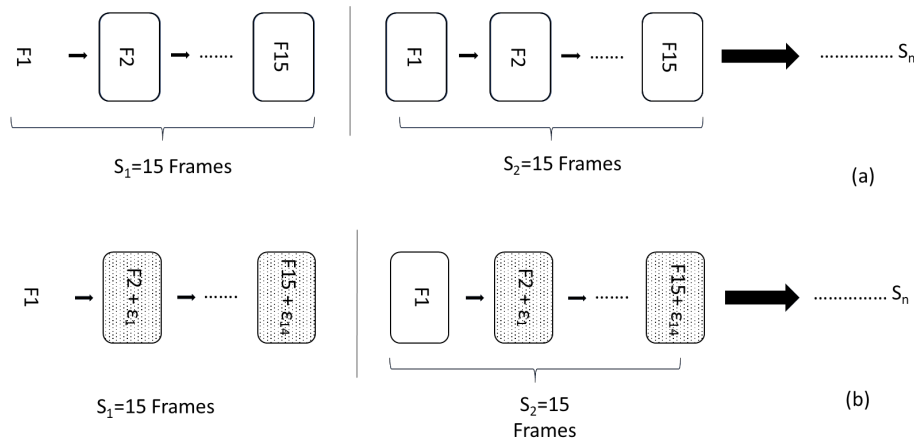
The top row in Figure 3 shows a change is detected and labeled as "Change Detected" at frame 16 as g_k crosses the threshold in the presence of Gaussian noise. The bottom row in Figure 3 shows g_k for the first 350 frames of the video. As expected, a spike in the decision function is observed every 15 frames, denoting a statistically significant change.

Figure 3

Temporal Change Detection. A Homogeneous Scene Has 15 Frames (1 Original + 14 Replicas)
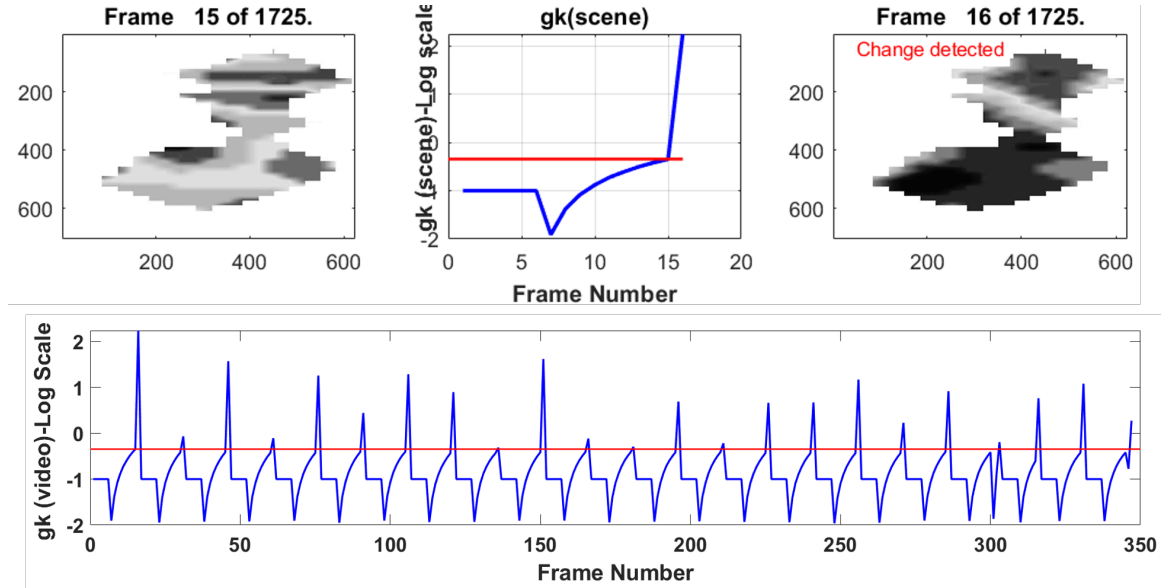
(Upper) form left to right shows frame 15 of the first scene, g_k in the presence of Gaussian noise with variance $\sigma^2 = 0.01$, and the first frame of the second scene with change detected; (Lower) depicts g_k for the first 350 frames. A spike in the decision function is observed every 15 frames, denoting a statistically significant change

Spatial Change Detection: To assess the spatial change detection, we used different sizes of macroblocks to estimate the features Y_k. The macroblock sizes tested were $M = 4 \times 4$, 8×8 and 16×16. Note that each pixel in the original image is equivalent to $900m^2 (30m \times 30m)$ in the land (the resolution of Landsat 8). Therefore, using higher macroblock sizes correspond to covering a larger land area and thus, may miss local temperature variations. For small macroblocks, we did not perform DCT and directly computed the PCA for each macroblock. We found experimentally that considering macroblocks of size 4 achieved the best detection for this dataset. The threshold h was set to balance the false positive and negative rates.

Figure 4

Spatial Change Detection of Bridgeton, MI, Landfill. The Area of Change Is Colored in Red for Images Dated on (a) January 27, 2000; (b) August 9, 2001; (c) January 24, 2005; (d) May 4, 2011
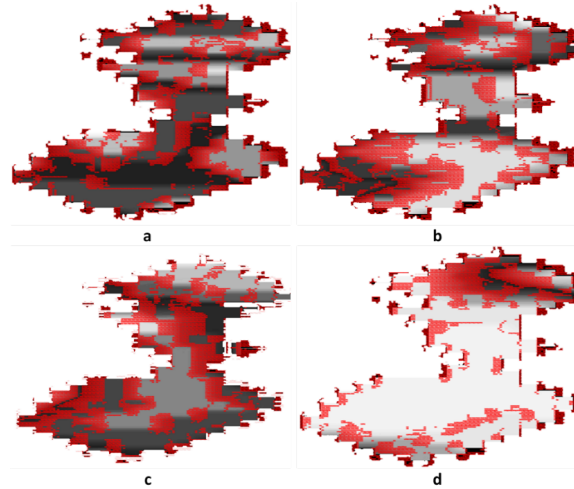
Figure 4(a-d) shows the spatial change detection of Bridgeton, MI, Landfill on different observations dates. The area of change is colored in red and obtained with macroblocks of size 4, threshold $h = 0.45$. We used 11 macroblocks to estimate θ_0. Notice that the algorithm correctly identifies the spatial regions of temperature change detection by observing the red colors that are plotted at the points where there is a change in temperature.

The interpretation of the publicly available Landsat data archive permits monitoring of large areas, such as landfills. The non-destructive, non-invasive methods described herein allow for observation of multiple locations quickly and at low cost, enabling assembly of a satellite image archive that shows changes in thermal state of landfill surfaces. In this chapter we introduced a new approach to the problem of change detection in LST remote sensing observations based on statistical sequential analysis theory. The reduced dimensionality aims at increasing the computational efficiency taking into account the large size of remote sensing data. The statistical approach is based on detecting the change in the mean parameter of the generative distribution of the stochastic data. In chapter 6, we

will consider data-driven approaches, based on deep learning models to detect changes in the data.

Chapter 5

Evaluating the Spatial Temperature Trends Using Clustering

First we will discuss the temporal temperature trends as it is the first sign of heat elevation in landfill and from there we will see the need for future spatial analysis. Temporal temperature trends is looking at the landfill as a whole system characterized by its basic statistics, such as the average and the minimum/maximum LST. Temporal temperature trends show when the temperature anomaly is happening and whether the trend is increasing or decreasing without specifying the location. The spatial analysis identifies the location of the hotspots, where these hotspots are developed and how they move, expand, or shrink over time. For the spatial analysis, we will show different algorithms. K-means clustering is an unsupervised clustering method based on numerical temperature values of LST. These results will be compared to the VAE K-means from LST images discussed in chapter 6. We also developed a Heat Index (HI) based on the deviation from the mean as a simplified clustering algorithm for comparison purposes as will be shown in chapter 7. The results have been considered for one active landfill in Deerfield Township, Cumberland County, New Jersey, USA and one closed landfill in South Harrison Township, Gloucester County, New Jersey, USA.

5.1 Temporal Behavior of the Landfill (Temporal Temperature Trends)

This simple temporal analysis is based on the difference between the highest and lowest LST for each observation, $\Delta LST = (LST_{max} - LST_{min})$. Under normal operating conditions, the landfill temperature remains close to the air temperature [28, 14] . On one hand, higher temperature differences could be an initial indication of subsurface smoldering event (SSE) or subsurface fire. On the other hand, this difference cannot tell us where the change is happening, rather it identifies a temperature anomaly that should be investi-

31

gated.

Figure 5a compares the LST_{min} and LST_{max} measured at the Deerfield landfill for the period of study and the difference between them. It shows a periodic cycle due to seasonal changes, whereby the LST is higher during the summer, late spring and lower during the winter or late fall. To reduce the seasonality effect and to show the general trend of temperature along the years, Figure 5a is redrawn in Figure 5b using a moving average of window size = 20 (smoothing the curve). In Figure 5a, the temperature maintained around the mean difference i.e., mean (ΔLST), (pink dashed line) until 2008, then LST reached its maximum at 28 °C. Figure 5b shows, in mid-2008 the LST difference consistently increased to be higher than the mean difference until the end of 2019 (except for the year between beginning of 2013 – beginning of 2014). Even though ΔLST is expected to increase somewhat as a normal result of subsurface processes in the landfill, especially after a few months of new waste is accepted [56], differences that systematically increase over time, which could be months or years, indicate changes in the subsurface activities such as a smoldering or fire event. As an example of a closed landfill, we show the results for the South Harrison landfill, which closed in 2012. Figure 5c depicts the LST_{min}, LST_{max}, mean(LST) and mean (ΔLST). The landfill shows normal operation during the years as the LST difference is kept around the mean. The effect of shutting down the landfill in 2012 started to appear in mid-2016 by the decrease in mean difference.

There are many factors that cause the temperature elevation in landfills. Organic and chemical wastes tend to generate more gases and yield higher temperature. The age of waste is also important. Generally, more recently accepted waste (after a few months up to a few years) produces more landfill gas through bacterial decomposition. Also, the water penetration due to rain will cause the cycle of gas production to restart again [56]. Hence, closed landfills must be monitored for subsurface activities that can lead to above-surface hazards. This can be shown in Figure 5c where the landfill continued to show

32

Figure 5

(a)Comparison of LST_{min} and LST_{max} Temperature at Deerfield; (b) Smoothing the Deerfield Curve Using Moving Average ($W = 20$); (c) Comparison LST_{min} and LST_{max} Temperature at South Harrison Using Smoothed Curves With Moving Average ($W = 20$)
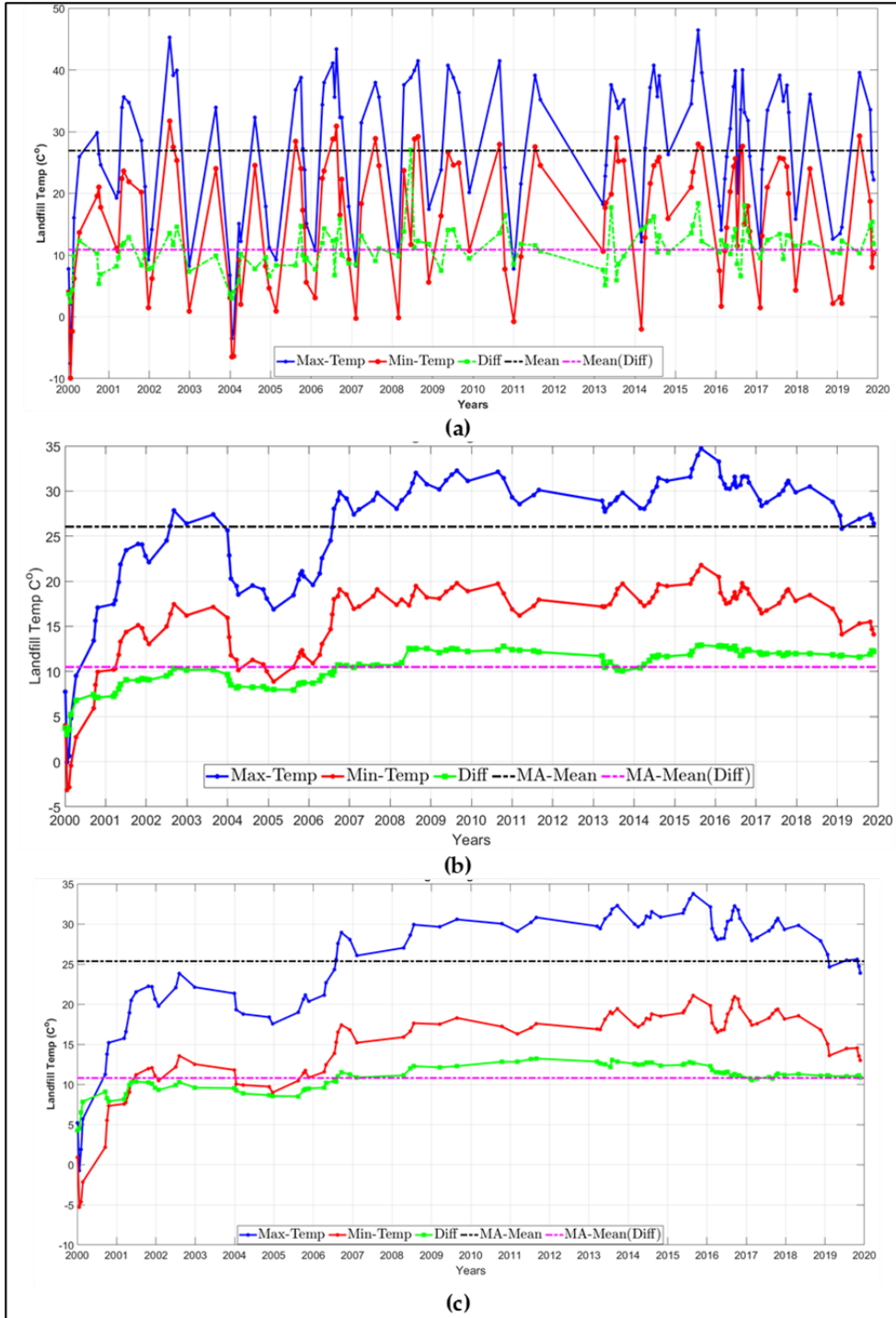
elevated temperature for a few years (2012-2016) after the closing date. Tables (A1-A10) in Appendix A shows multi-temporal maps (thermal maps) for the landfills listed in Table 1 in which the hotspots or the anomaly areas are easy to discern based on the LST.

Thus far, we have shown that, by examining the temporal behavior of the landfill, we were able to detect temperature anomalies as in Figure 5 from satellite images. Next, we show temperature changes in the spatial domain and trace the movement of the hotspots within the landfill during the years of the study.

5.2 Spatial Behavior of the Landfill at Pixel Level

As mentioned above, knowing the maximum LST measured is not enough to locate the hotspots and track their movement within the landfill. To show the thermal state and anomaly area within the landfill in the spatial domain during the 20 years of study, we used K-means clustering for the LST temperature values(TIF files) and to compared it with VAE K-means clustering in chapter 6.

5.2.1 K-Means Clustering

In chapter 1, we have discussed K-means algorithm as an unsupervised clustering method, and how it can be used for unsupervised anomaly detection. Generally, the output of the anomaly detection algorithms is a binary label given to every data point to indicate normal or anomaly. However, in multiple clustering ($k > 2$), it is application dependent and it is for the analyst to define the anomaly cluster(s) by given scores for each cluster [7]. In this work, we used hard clustering using K-means algorithm with $k = 4$. For the purpose of tracking the heat elevation and anomaly areas in the landfill, we used $k = 4$ to cluster each LST observation to four clusters, with cluster 1 grouping the lowest temperature and cluster 4 grouping the highest temperature. The four clusters are denoted as the following:

- Cluster 1 represents No Risk.

34

- Cluster 2 represents Lower Risk.

- Cluster 3 represents Moderate Risk.

- Cluster 4 represents Higher Risk.

LST Temperature normalization, changes the values to a common scale, that make observations comparable either in the same landfill or at different landfills. Consequently, the clustering results (1-4) will be comparable between different observation.

Figure 6

K-Means Clustering Results for LST Observation on July 19, 2013 Deerfield Landfill
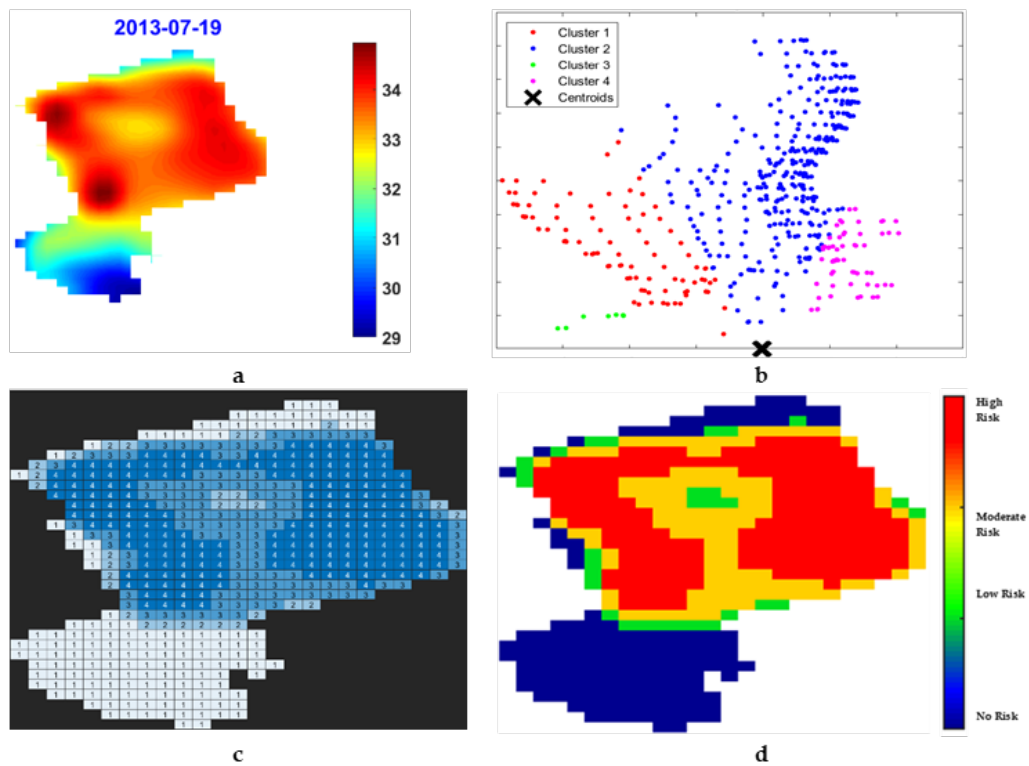
(a) LST on July 19, 2013 Deerfield Landfill, NJ; (b) K-means clustering results; (c) tracing cluster numbers(1-4) back to their respective pixels; (d) plotting the clusters in color to denote their risk level

35

Figure 7

Developing of Hotspots in Deerfield Landfill, NJ, From February to August 2014

(upper row) shows the calculated LST observation in $^{\circ}C$ on that day; (lower row) shows the K-means clustering for the corresponded LST observation.

In Figure 7, we show the K-means clustering algorithm results form 6 consecutive observations from February 2014 until August 2014. Row 1 shows the calculated LST and row 2 shows the corresponded K-means clustering with $k = 4$, grouping each LST observation into 4 clusters, with cluster 1 grouping the lowest temperature (no risk in blue color) and cluster 4 grou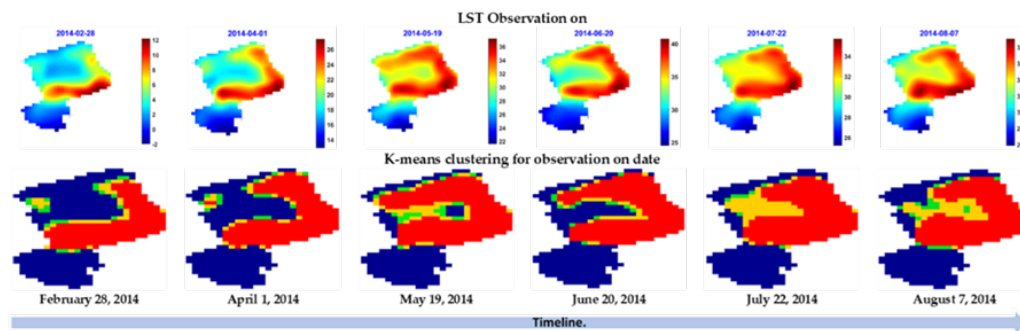ping the highest temperature (higher risk in red color). Row1 shows continuous heat elevation expanding from February to August results in hotspots expanding from the east of the landfill to the west. Figure 7 row 2, clearly depicts that as the temperature continues to rise in the landfill, the no risk (blue) area at the East of the landfill starts to shrink and is replaced by higher temperature shown in green or orange colors. Eventually, by August 2014 the upper side of the landfill demonstrates temperature elevation ranging from moderate to higher risk and the no risk area is decreasing continuously.

Figure 7 explains the effectiveness of using K-means clustering as an unsupervised clustering algorithm in detecting hotspots development and movement across the landfill in the absence of data from landfill operators. The links for video files that show the movement and development of hotspots are given in Appendix A for all landfills in Table 1.

5.2.2 *Heat Index (HI) and Accumulated Heat Index (AHI)*

We developed a Heat Index to further investigate the behavior of the landfill spatially. It should be noted here that the term heat index has nothing to do with any thermodynamics terminologies, it is a number to classify every pixel in the landfill according to its deviation from the mean. Heat index (HI) is used to give an index to every pixel in each observation at a given date. Since we are concerned with heat elevation, all observation below the mean were assigned an index (-1) to represent no risk, other temperatures are assigned 1, 2, ... etc, depending on its standard deviation interval as shown in Figure 8 (left block).

Figure 8

Block diagram of Algorithm 2 calculating Heat Index (HI) and Accumulated Heat Index (AHI)
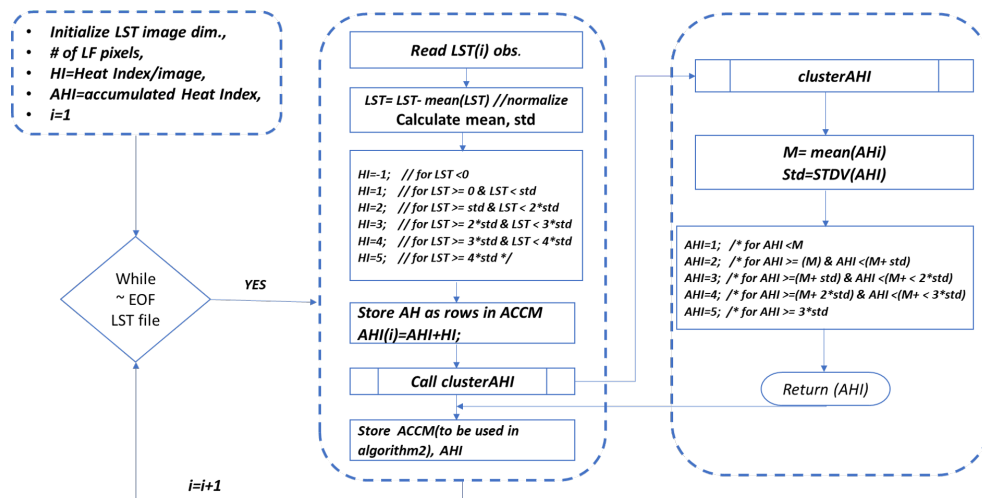
(a) *HI* is a spatial analysis used to give an index to every pixel in each observation based on its deviation from the mean; (b) Accumulated Heat Index (AHI) is used for the spatio-temporal analysis to show the average of landfill indices. It is obtained by sequentially summing LST observations then averaging and re-indexing. *AHI* is a matrix ($l \times w$) of the landfill size, where l and w are the length and width of the landfill respectively; (c) *ACCM* is a matrix ($n \times m$), where n is the number of pixels of the landfill and m is the number of observations. It is used to store HI row wise for later use by Algorithm 3.

All HI indices for one observation are stored row wise in *ACCM* to be used later by Algorithm 3 to quantify the landfill health state. *ACCM* is a matrix ($n \times m$), where n is the number of pixels in the landfill and m is the number of observations. Tracing any column, will show the history of landfill indices from the beginning of the study in year 2000.

Algorithm 2: Indices to quantify landfill Health state

Input: $AHI(l,w) := 0$, $ACCM(n,m) := 0$

    /* where l, w are the length and the width of the landfill and   */

    /* n, m are the #pixels of the landfill and m is #observation   */

1   $i := i$

2   **while** $\sim EOF\ (LST)$ **do**

3       $Read\ LST(i)$

4       $LST = LST - mean(LST)$ // normalize

5       $\sigma := std(LST)$

6       $HI := -1$ // For $LST < 0$

7       $HI := 1$ // For $LST(\geq\ 0\ \&\ < \sigma)$

8       $HI := 2$ // for $LST(\geq \sigma\ \&\ < 2\sigma)$

9       $HI := 3$ // for $LST(\geq 2\sigma\ \&\ < 3\sigma)$

10      $HI := 4$ // for $LST(\geq 3\sigma)$

11      $Reshape(HI(1,:)) \rightarrow ACCM$ // reshape *HI* and save as a row in *ACCM*

12      $AHI := AHI + HI$ // accumulate *HI* in *AHI*

13      $CALL\ Function(ClusterAHI)$ // function to average, re-cluster and print image

14      $Save\ AH, ACCM$ // store to be used by Algorithm 3

15      $i := i + 1$

16   **return** (Plots as in Figure 21 and Figure 22)

17 **Function** ClusterAHI(*AHI*):

18      $AHItemp := AHI$

19      $\mu := mean(AHItemp)$, $\sigma := std(AHItemp)$

      /* Re-cluster accumulated observation   */

20      $AHItemp := 1$ // $AHItemp <\ \mu$

21      $AHItemp := 2$ // $AHItemp(\geq \mu\ \&\ < \sigma)$

22      $AHItemp := 3$ // $AHItemp(\geq \sigma\ \&\ < 2\sigma)$

23      $AHItemp := 4$ // $AHItemp(\geq 2\sigma\ \&\ < 3\sigma)$

24      $AHItemp := 5$ // $AHItemp \geq 3\sigma)$

25   **return** (Print average indices for accumulated observations)

5.3 Spatio-Temporal Behavior of the Landfill

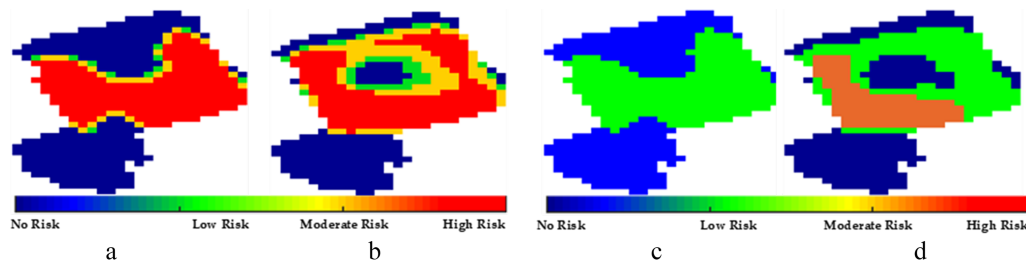### 5.3.1 Accumulated Heat Index (AHI)

The accumulated heat index (*AHI*) is used for the spatio-temporal analysis to show the history of landfill indices. It is obtained by summing LST observations as they are read as explained in Algorithm 2 Figure 8 (left block). *AHI* is a matrix ($l \times w$) of the landfill size, where *l* and *w* are the length and width of the landfill respectively. Averaging and re-indexing the content of *AHI* at any give date shows the average clustering for all the previous observation until that date. It is also stored to be used with Algorithm 3 as will be explained in chapter 8

To further explain the accumulated indices, Figure 9a shows the K-mean clustering on 25-11-2019, i.e. the thermal state of the landfill on this date. Figure 9b shows the accumulated K-means clustering until 25-11-2019, which is obtained by averaging all the previous LST indices starting from the first observation and re-clustering again until that date. This approach will show continuously elevated temperature areas with high cluster numbers 3 or 4 and moderate risk (orange) or higher risk (red) respectively. The areas maintaining low temperatures will be assigned a small cluster numbers 1 or 2 and no risk (blue) or lower risk (green) respectively. In Figure 9b, red color denotes an area that has a long history of heat elevation and probably a remedy actions should be taken. On contrary, blue color denotes no risk area. From a few personal visit to Cumberland county – Deerfield landfill- we know that the South area of the landfill was opened to accept dump only after March 2016, which could be a reason why it maintains -no risk area- for the whole duration of study. Again in Figure 9b, the orange color area denotes -moderate risk- and has to be monitored for the possible escalation to a higher risk area if heat elevations continued in the following observations.

39

Both K-means and HI Figure 9 (c, d) and their respective accumulated indices have delineated the same area. However, accumulated K-means and accumulated heat index differ slightly in defining hotspots until 25-11-2019. While accumulated K-means is more specific in defining hotspots and the elevated temperature areas around them, the accumulated heat index is more conservative in defining hotspots, as shown Figure 9 (c, d). The colors (blue, green, orange and red) corresponding to (no risk, lower risk, moderate risk and higher risk) areas, respectively, and reflect the average thermal state of the landfill at this date.

Figure 9

Comparing K-Means Clustering and Heat Index (HI)

(a,b) K-means and accumulated K-mean clustering for Deerfield landfill on /until 25-11-2019 respectively; (c,d) Heat index and accumulated heat index (AHI) for Deerfield landfill on /until 25-11-2019 respectively.
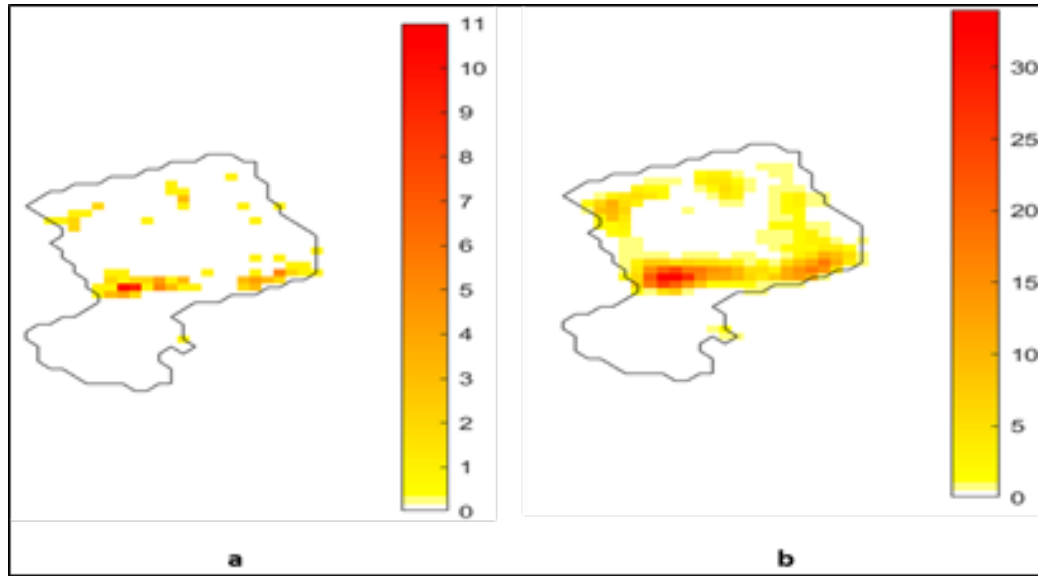
5.3.2   *Frequency of Maxima (FM)/Frequency of Near Maxima (FNM)*

Frequency of maxima is a measure of the number of times a given spot in the landfill has recorded the highest temperature during the period of study. The frequency of near maxima is a measure of the number of times a given spot in the landfill has recorded the highest temperature or near highest temperature (within -0.5 °C from the maximum temperature). The bar on the right side of both figures in Figure 10 (a, b) indicates the frequency of occurrence of maximum/near maximum temperature at that point. This analysis, helps to predict in which area in the landfill an internal fire is more likely to occur. Comparing

Figure 10, we see that FNM gives results that are very close to the accumulated heat index in Figure 9 (b, d).

Figure 10

The Number of Times a Given Spot in Deerfield Landfill Has Recorded

(a) The highest temperature (frequency of maxima); (b) near the highest temperature (frequency of near maxima)

Chapter 6

Application of Remote Sensing and Deep Learning in Detecting Internal

Temperature Anomalies in Landfills

The spatial analysis of the landfill identifies the location of the hotspots, where these hotspots are developed and how they move, expand, or shrink along the time. For the spatial analysis we show different algorithms based on state-of-the-art unsupervised deep learning method of VAE to detect temperature anomalies in landfills based on both feature extraction of the encoder module and reconstruction error of the model. The former, is to use the encoder module of the VAE to extract low-dimensional features of the image and feed them to a K-means clustering algorithm (VAE K-means). VAE K-means is used to either cluster the thermal status of the landfill with $K = 4$ or detect anomaly areas with $K = 2$. The latter, train a VAE to learn the distribution of normal data (without anomaly). When a new data with anomaly is fed to the model, the anomaly areas can be identified and localized using the reconstruction error. Then we compare the results of both to the classical K-means clustering for the raster data (temperature values).

## 6.1 Variational Auto-Encoders (VAE) Overview

VAE is a convolutional neural network (CNN), which is trained to reconstruct the input data. It has two main parts; an encoder that compresses the input data to a low-dimensional latent space, and a decoder to reconstruct the input data from the latent space.The encoder outputs two vectors describing the parameters of the distribution, the mean and variance (assuming Gaussian distribution). The decoder generates a latent vector by sampling from the mean and variance vectors to reconstruct the input image as shown in Figure 11. During the training process, the VAE learns to minimize the reconstruction error and to extract the salient feature of the input image. After training, the latent space

provides a (non-linear) lower-dimensional representation of the input data. Figure 12 shows VAE as a probabilistic generative model, the encoder is a probabilistic model and is given by the function $q_\phi(z|x)$, where ϕ is the learnable parameters of the encoder that need to be optimized through back propagation. The distribution $q_\phi(z|x)$ is assumed to be a good approximation to the posterior distribution $P_\theta(z|x)$. Latent vector z is formed by sampling from the mean (μ) vector and variance (σ^2) vector that is representing the diagonal elements of the covariance matrix. The decoder is also a probabilistic model and it is given by the $p_\theta(x|z)$ that defines the likelihood of input x given z, where θ is the learnable parameters of the decoder that need to be optimized through back propagation.

Figure 12 also depicts the structure of VAE where input image x is passed to the encoder $q_\phi(z|x)$ of VAE to output the parameters that describe the distribution, the mean and variance vectors, then latent space z is sampled from the mean and variance. The decoder $p_\theta(x|z)$ learns to reconstruct, i.e., to generate the original data x from z.

Figure 11

Variational AutoEncoder (VAE) as a Probabilistic Generative Model
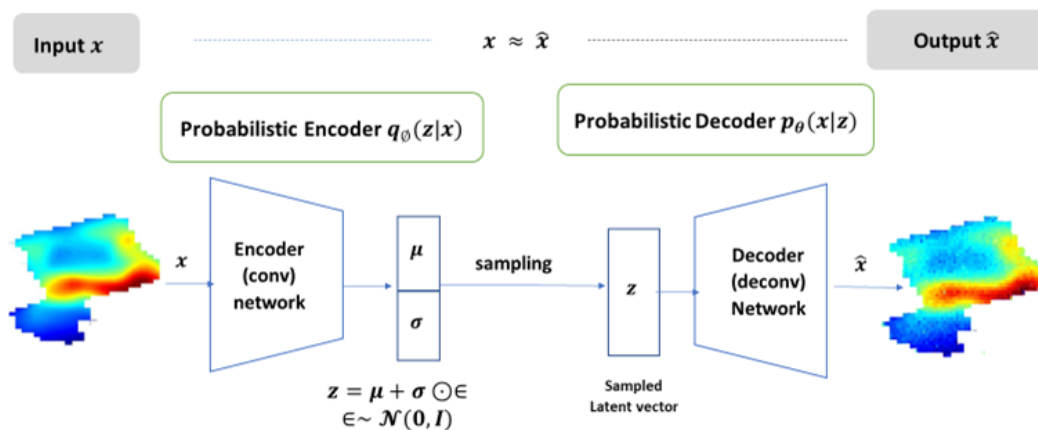
The encoder $q_\phi(z|x)$ compresses the data into a smaller dimension, which is then fed to the decoder $p_\theta(x|z)$ to reconstruct the input image with minimum error using back propagation. During back propagation, the value of z is replaced by the values shown at the bottom in order to allow for back propagation through a deterministic node instead of stochastic distribution (reparameterization trick).

43

From the discussion above, an encoder-decoder networks can be thought of as a conditional probability. The encoder is the conditional probability $q_\emptyset(z|x)$ of generating the latent vector z given the input data x and it is trained to approximate the true posterior distribution $P_\theta(z|x)$. The decoder is a conditional probability $p_\theta(x|z)$ of reconstructing the original input given the latent vector z and it is trained to learn likelihood distribution of data. Both $\emptyset$ and θ are the tunable parameters (weights and bias) for the encoder and decoder respectively that need to be learned. The loss function of VAE has two terms,

$$VAE_{Loss} = Reconstruction\ loss + Regularization, \tag{17}$$

$$VAE_{Loss} = ||x - \hat{x}||^2 + KL[\mathcal{N}(\mu_\theta, \sigma_\theta), \mathcal{N}(0,1)], \tag{18}$$

Where x is the input image and $\hat{x}$ is the reconstructed image. Following the derivation in [57], the loss function can be written as:

$$VAE_{Loss} = E_{q_\emptyset(z|x)}[log\ p_\theta(x|z)] - KL[q_\emptyset(z|x)||p(z)]. \tag{19}$$

Where, KL is the Kullback-Leibler divergence that measures the difference between the two distributions $q_\emptyset(z|x)$ and P(z), z is the latent space, x is the input data and $p(z) = \mathcal{N}(0,1)$. During training, VAE minimizes the loss, as can be seen from Equation 18 and Equation 19.

44

Figure 12

The Structure of VAE.
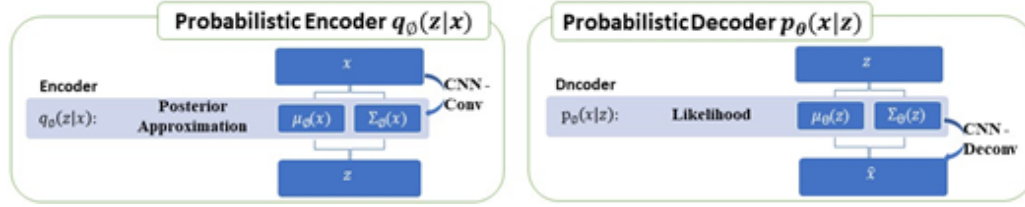
Input image *x* is fed to the encoder $q_\emptyset(z|x)$ through the CCN (convolution) network to output the parameters that describe the distribution, the mean and variance vectors, then latent space *z* is sampled from the mean and variance. The decoder $p_\theta(x|z)$ learns to generate the original data *x* from *z* (deconvolution).

The network parameters $\emptyset$ and θ in Equation 19 and Figure 12 are optimized during training using back propagation of the gradient of the loss. Since in forward propagation the latent vector *z* is formed by sampling from the mean (μ) and variance (σ) vectors to form a Gaussian distribution, it is not possible to back propagate though a stochastic distribution to adjust the weights. Therefore, a reparameterization trick is used to replace the value of *z* by the values shown at the bottom of Figure 11 to allow for back propagation through a deterministic node instead of a stochastic distribution. The value of *z* will be μ added to standard Gaussian distribution scaled by σ ($z = \mu + \sigma \odot \varepsilon$ where $\varepsilon \sim \mathcal{N}(0,1)$ and $\odot$ is element-wise multiplication operator. The full derivation of VAE loss function and the reparameterization trick has been described by several researchers [57, 58, 59].

6.2 VAE K-Means Clustering

The use of VAE as described above, considering its ability to extract a reduced dimensionality of the input image while it is still maintaining the important image's information, makes it suitable for LST observation images and for remote sensing data in general. The nature of heat elevation in landfills dictates that multiple anomaly areas can exist in the same observation; hence, the detection problem should consider every observation in small patches of reasonable size. Therefore, for a given LST image *x* fed to encoder

45

model, x is divided into patches P_n, where n is the patch number such that $n \in [1:N]$ and N is the total number of patches. Then each P_n is fed to the encoder model to generate a useful representation of the patch in its latent space vector L.

The workflow of VAE operation is summarized in Figure 13. For this application, the aim is to extract features for every patch using dataset-1. Hence, each LST observation is divided into patches of size 16×16 and the VAE will extract 3600 feature vectors. These features are grouped to pass to the K-means algorithm to classify the thermal status of the landfill during the test phase. An 8×8 patch size was tested, producing 14400 feature vectors with a lower reconstruction error, but it taking four times the processing time.

Figure 13

VAE Schematic for Feature Extraction From an Image, Then Passed to a K-Means Algorithm for Clustering
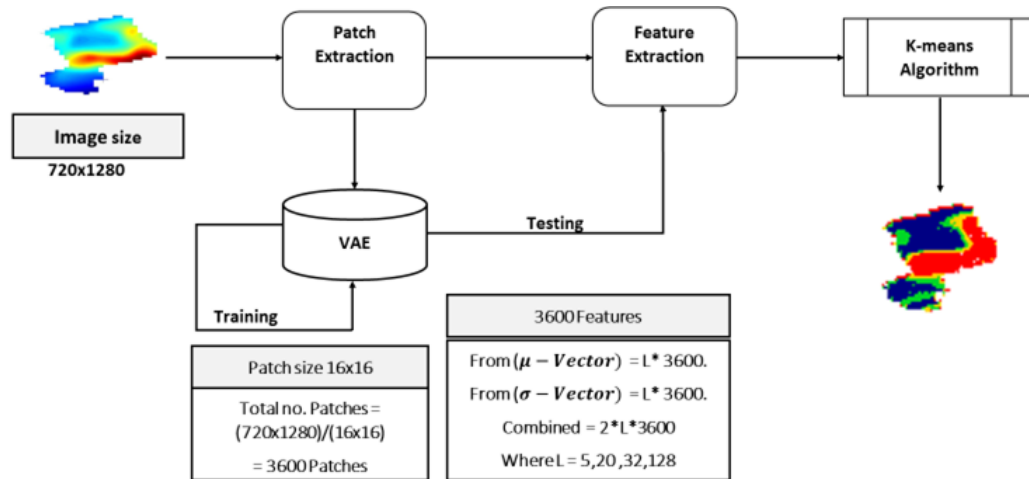

Figure 7 in chapter 5 is redrawn in Figure 14 to compare the results of the classical K-means results to the VAE K-means results. Figure 14 depicts how a hotspot in Deerfield Landfill, NJ developed and expanded from February to August 2014. Figure 14 the upper row shows individual LST observations measured in degrees Celsius obtained from

calculating LST from satellite images between February 28, 2014 to August 17, 2014. In the second row the K-mean clustering with $K = 4$ for each individual observation in those dates based on temperature values in $^{\circ}C$. Row 3 and 4 are the results obtained from extracting features of LST images (in PNG format) by the trained VAE after being fed to the VAE K-means algorithm for $K = 4$ and $K = 2$, respectively. It is clearly shown that the use of reduced dimensionality obtained from the latent space during testing can give similar clustering for $K = 4$. The fourth row for $K = 2$ can detect the anomaly temperature in the landfill.

A high variation in LST is most of the time caused by subsurface heat activities such as SSE or subsurface fire. Subsurface smoldering events can go unnoticed for a long time and continue to cause high temperature difference on the surface. For instance, if there is SSE in one area in a landfill that lasts for a long time it will be detected by satellite images and appear in the LST observation for several consecutive observations. The analysis can show that area is persistently high compared to other areas in the landfill as we read more observations.

Figure 14

Tracing Hotspots in Deerfield Landfill, NJ From February to August 2014
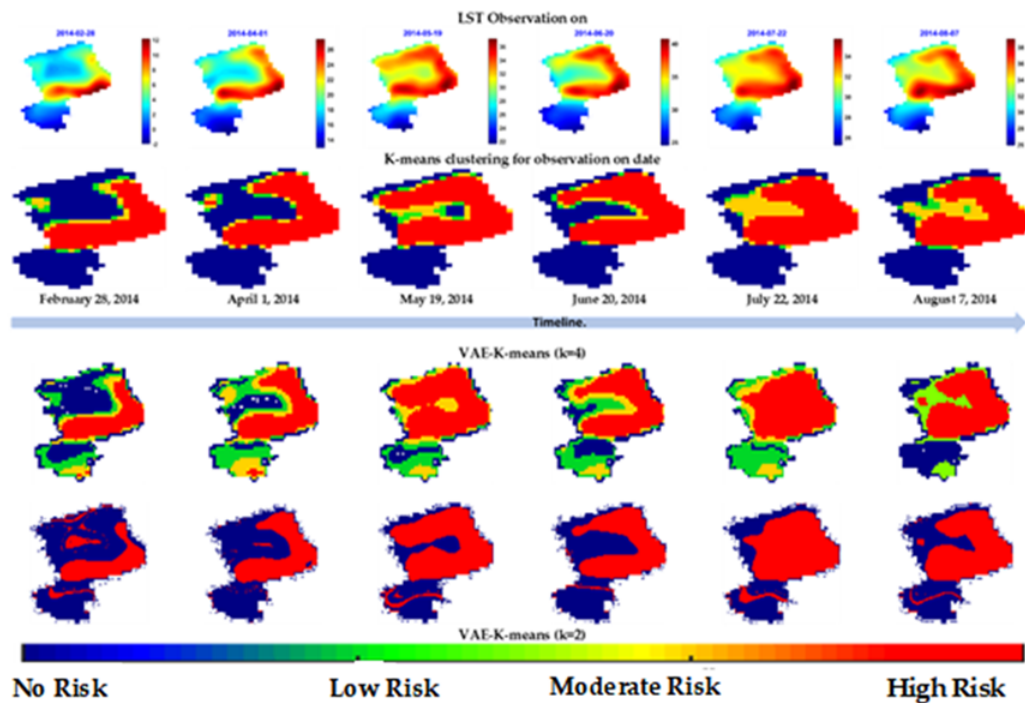
(second row) shows the thermal status of the landfill using classic K-means clustering with $K = 4$ for each individual LST observation on that day;(third row) shows the thermal status of the landfill using the extracted features of VAE K-means with $K = 4$;(lower row) shows the anomaly temperature in the landfill using $K = 2$.

6.3 Variational Auto-Encoder for Anomaly Detection and Localization Based on Reconstruction Error

VAE can also be used to detect and localize the anomaly areas in landfills based on the reconstruction error. For this purpose, we used the customized dataset-2 as described in chapter 3 and trained the network to learn the distributions of a normal dataset (without anomalies). During testing, the network will not be able to reconstruct the parts of the images with anomalies areas. The difference between the input and the output of the VAE is clearly identified and locates the anomaly areas in the landfill. Figure 15 shows a trained VAE with an image that has a wide anomaly area in the input x and in the output $\hat{x}$ is the reconstructed image. The difference between them is a gray-scale image showing the

48

anomaly area with some reconstruction error. Converting the difference to a binary image

shows the anomaly and its location in the landfill.

Figure 15

VAE to Detect and Localize the Anomaly Areas in Landfills
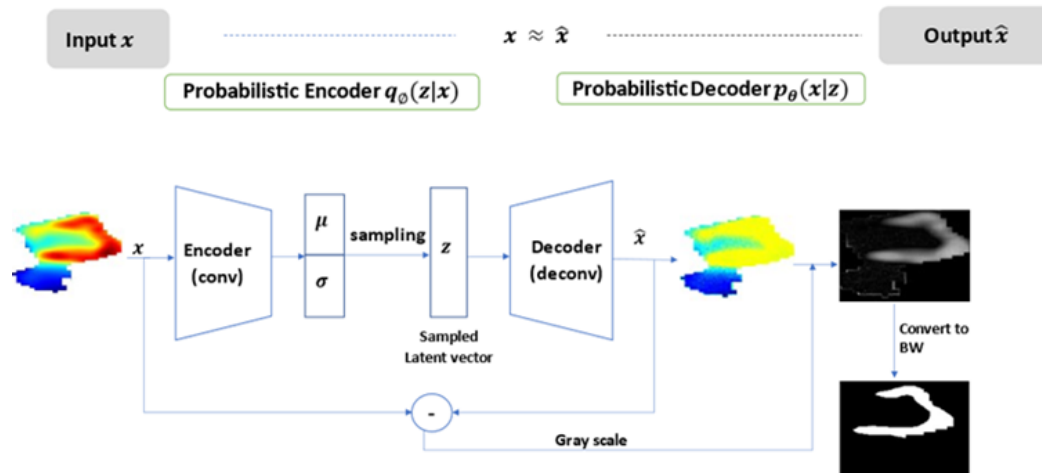
The input image x is fed to VAE; the decoder output $\hat{x}$ cant reconstruct the red areas (anomaly); the difference $(x - \hat{x})$ shows and locate the anomaly area. The small error is eliminated by converting the image to black and white image. Larger error is eliminated using simple morphological operations.

The anomaly detection using VAE in Figure 15 is based on a simple idea that is, if

it is not normal then it is anomaly. Therefore, it should be noted that the accuracy of this

method greatly depends on the definition of normal images.

Figure 16

VAE to Detect and Localize the Anomaly Areas in Landfills
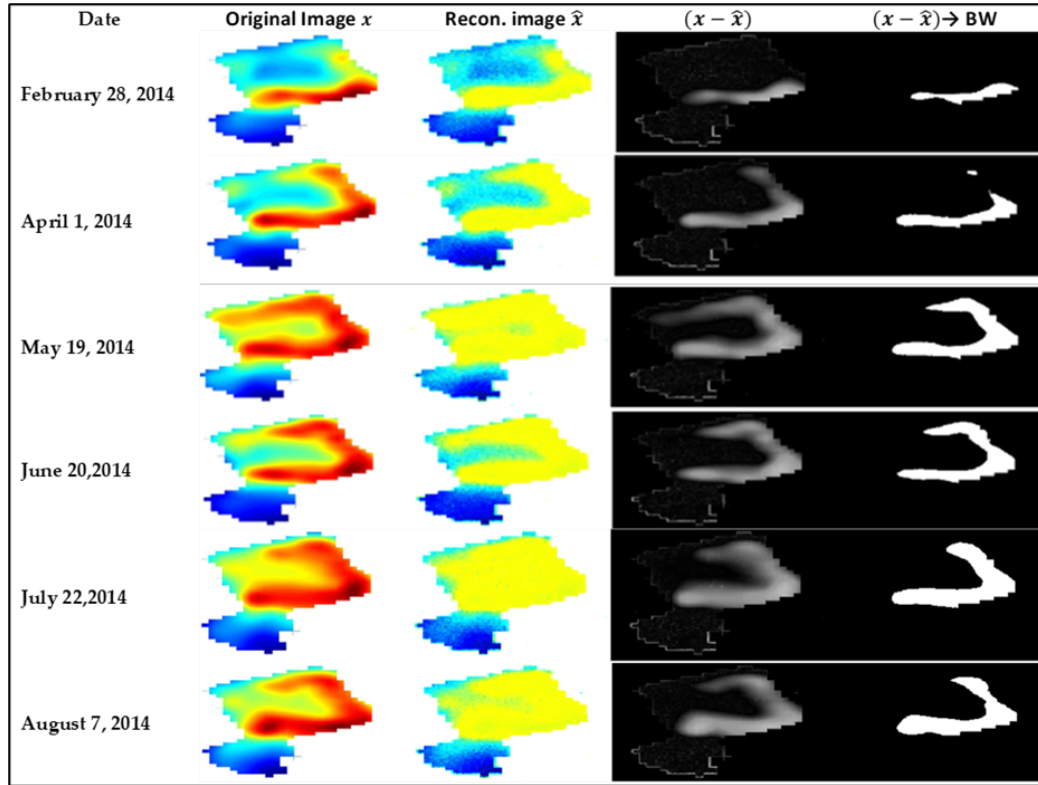

Row wise, at any given date, the input image *x* is fed to VAE; the decoder output $\hat{x}$ cant reconstruct the red areas (anomaly); the difference $(x - \hat{x})$ shows and locate the anomaly area, The small error is eliminated by converting the image to black and white image with threshold = .5, larger error is eliminated using higher threshold or simple morphological operations.

The results shown in Figure 16 are obtained by training the VAE with customized dataset-2 that has no or few anomaly images. During testing using the images (February 28, 2014 to August 17, 2014), the decoder will not be able to reconstruct the anomaly parts of the images as it is not trained to see it. Therefore, the difference $(x - \hat{x})$ between the input image *x* and the reconstructed image $\hat{x}$ shows the anomaly area. The difference is shown in gray-scale image with some reconstruction error. The small errors can be eliminated by converting the image to a binary format with threshold = .5, larger errors can be be eliminated using higher thresholds or simple morphological operations, if needed. The results of Figure 16 can be compared to that of Figure 14 rows 4 where the same area of

anomaly is detected. All methods used were able to identify the same area of anomaly

Comparing VAE K-means to VAE for anomaly detection and localization, we found that VAE for anomaly detection is more accurate in identifying and isolating the anomaly area as in Figure 16. It does not require a large dataset. It can be performed with as low as 40-60 training images. However, the dataset used consists of 110 images (90 for training and 20 for validation). Since VAE for anomaly detection depends only on the difference between input and output to detect the anomaly area, it is much faster in getting the results. On the other hand, the VAE K-means is time consuming. We found that the feature extraction is very much dependent on the patch size and the dimensions of the latent space variable. As explained in Figure 13c, reducing the patch size form 16x16 to 8x8 will increase the number of features, but the processing time is increased by 4 fold. The accuracy of VAE K-Means depends on patch size, the smaller the patch size the lower is the reconstruction error.

The original data (LST observations) is high dimensional and contains redundant information. Thus, to extract salient features, VAE is used as a (non-linear) dimensionality reduction tool by inputting the image to a stochastic distribution of the latent space, from which the latent space variable is sampled from normal distribution, then the decoder will try to recontact the input image from latent space variable (low dimension representation of the original input) with minimum error through iterative training process. After training, the latent space learns to keeps the valuable information of the data with less or no redundancy. Yao et al., In [42] reported that some classical unsupervised methods such as Local Outlier Factor (LOF) and Cluster-Based Local Outlier Factor (CBLOF), could perform better with extracted features from VAE.

51

Chapter 7

Internal Clustering Validation and Determining Optimal Number of Clusters

Unlike supervised learning, where there is a ground truth to evaluate the model's performance, where the output data are labeled; clustering analysis, as mentioned in Section 1.1, relies on unlabeled datasets to learn and discover patterns in the data. Furthermore, because K-means takes K as input and does not learn it from data, there is no one correct answer in terms of the number of clusters that should be present in every issue.

One method to evaluate clustering findings is to begin by inspecting the clusters generated and making a decision based on our knowledge of what the data represents, what a cluster represents, and what the clustering is meant to achieve. However, there are a variety of quantitative methods for examining clustering findings that can be used to assess the quality of the clustering results.

Cluster validation is a technique for assessing the quality of clustering algorithm results. Internal cluster validation, which evaluates the quality of a clustering structure using just the internal information of the clustering process and without referring to external information. In general, it may be used to estimate the number of clusters and the best clustering technique in the absence of any external data.

The basic objective of clustering algorithms is to divide the dataset into groups of observations so that observations in the same cluster are as similar as possible and observations that are dissimilar are in different clusters. As a result, the internal clustering validation metrics represent both intra-cluster coherence and inter-cluster separation.

As mentioned in chapter 1, the lack of ground truth data is unfortunate. Therefore it is necessary to find some kind of metric to impose upon the output of the clustering algorithm that can shed light on the final clustering results. For this purpose, we use two of the most commonly used methods to determine the optimal number of clusters, Elbow and Silhouette coefficient methods.

7.1 Elbow Method

The elbow method calculates and plots the Within-Cluster-Sum of Squares (WCSS) for different values of K, which is the most often used method for determining the optimal number of clusters. The WCSS score is calculated as the sum of the square distances between each observation and the cluster center.

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} (\mathbf{x}_i - \mu_i)^2 \tag{20}$$

where $(\mathbf{x}_i - \mu_i)^2$, is the square distance between the ith observation and the cluster center μ_i and C_i is the number of observation assigned to cluster i.

The K for which WCSS shows a change from steep to shallow (appears like an elbow in the curve) will determine the optimal number of clusters k.

Elbow point can be found using the percentage of various explained, which can be calculated by Equation 21

$$\% \, Variance \, Explained = \frac{TSS - WCSS}{TSS} \times 100 \tag{21}$$

where TSS is the total sum of squares between every data point to tall data points.

$$TSS = \sum_{i} \sum_{j} (\mathbf{x}_i - \mathbf{x}_j)^2 \tag{22}$$

53

where TSS is the total sum of squares between every data point to tall data points. Similarly, the elbow point is the point where the relative improvement is not very high anymore.

7.1.1 Elbow Analysis for Conventional K-means Clustering

Recall Figure 14 second row in chapter 6, where we showed the results of k-mean clustering using K=4 for observation from February to August 2014 based on temperature values in $°C$. Figure 17, shows the corresponded elbow curves for the above images. Left column depicts the optimal number of clusters K obtained from elbow analysis based on WCSS. The right column is the elbow analysis based on the percentage of variance explained, which is the ratio of variance within the cluster to the total variance. The optimal cluster number k is taken at the corresponding 90% of variance explained.

Comparing Figure 17 (left column) and Figure 18, it is notable that for K-means WCSS range is approximately 500-2000, whereas the VAE K-means ranges approximately from 50 to 200. This indicates that the salient features extracted from the VAE encoder output tend to form compact clusters with minimum within cluster sum squared error (WCSS) compared to the conventional K-means.

7.1.2 Elbow Analysis for VAE K-means Clustering

Figure 18 shows the elbow analysis for VAE K-means for the images in Figure 14 row 3, the elbow curves obtained at percentage of variation at 95% and shows that the optimal cluster number is $K = 4$. However at 90% variation the optimal cluster number is $K = 3$ as shown in Table 3.
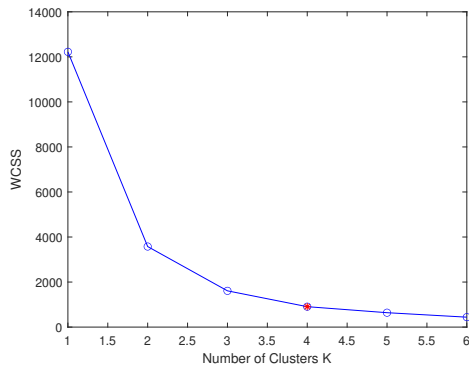
Figure 17

The Elbow Analysis for Observation Form February to August 2014 (Figure 14 row 2). Red Stars Indicate the Optimal Number of Clusters k for These Observations
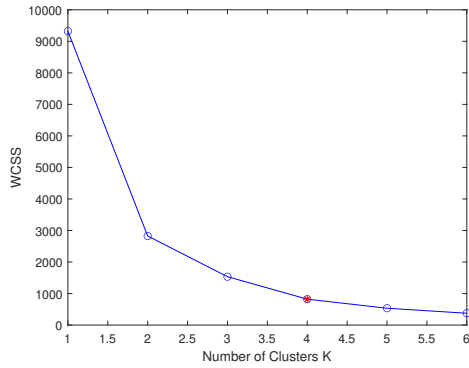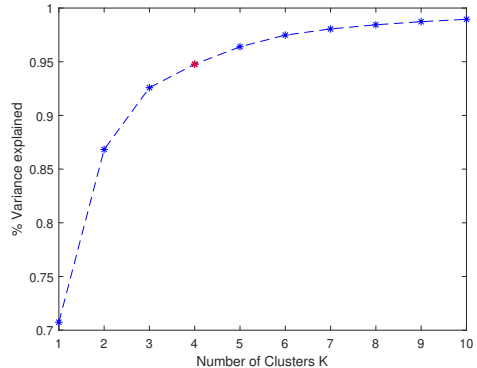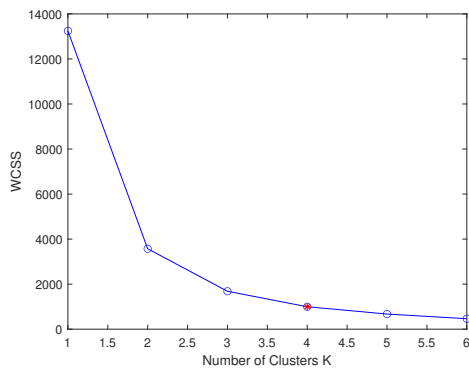

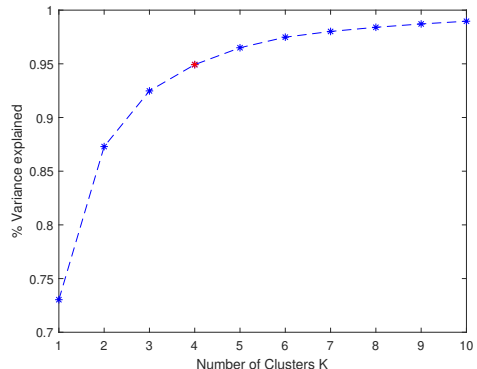**(a)** *February 28, 2014*
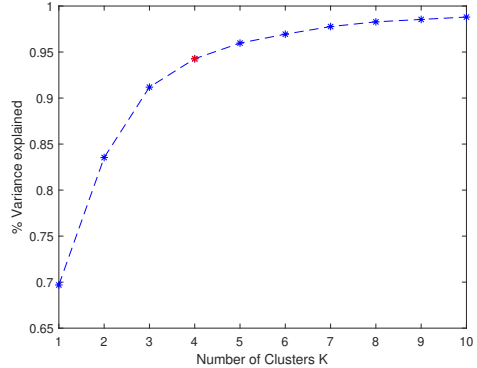
(b) *February 28, 2014*

(c) *April 1, 2014*

(d) *April 1, 2014*

(e) *May 19, 2014*

(f) *May 19, 2014*

55
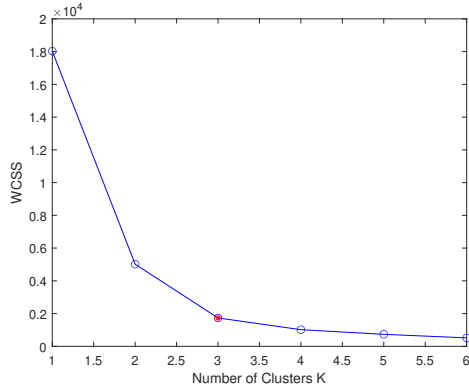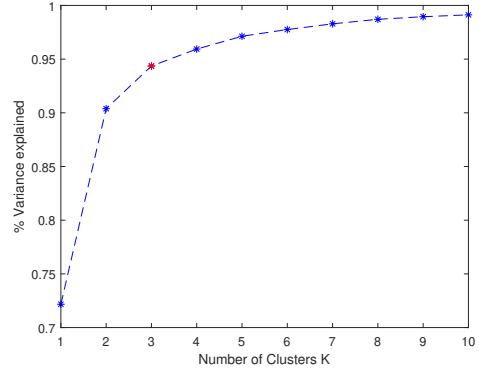
(g) *June 20, 2014*

(h) *June 20, 2014*
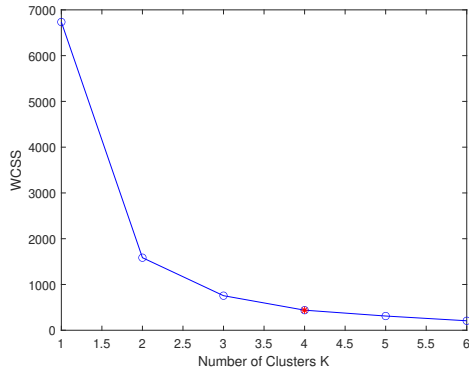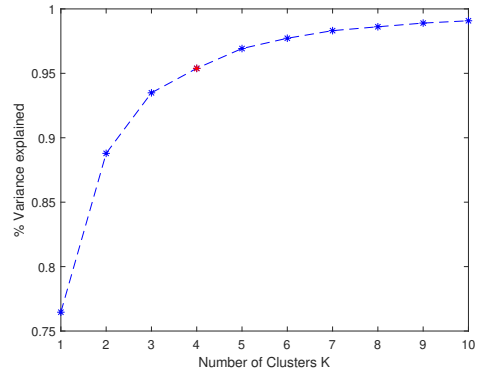
(i) *July 22, 2014*

(j) *July 22, 2014*
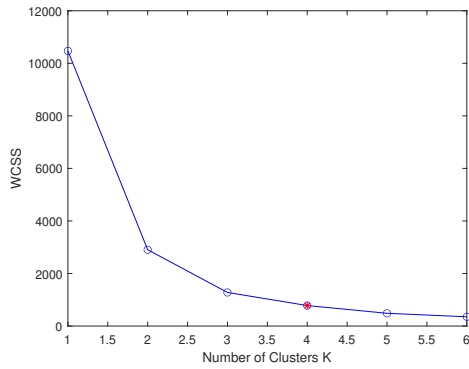
(k) *August 7, 2014*
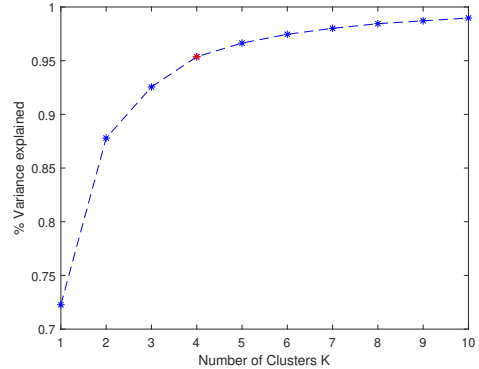
(l) *August 7, 2014*

Figures (a, c, e, g, i and k) are the elbow analysis based on minimum WCSS. Figures(b, d, f, h, j and l) are the elbow analysis is based on percent of variance explained, which is the ratio of the between-group variance to total variance.

7.2   Silhouette Method

The silhouette is another internal measure for cluster validation, it measures the

cohesion, which is the measure of how close the observations are within the same cluster.

56

Figure 18

The Elbow Analysis Based on Minimum WCSS for Observation Form February to August 2014 (Figure 14 Row 2). Red Stars Indicate the Optimal Number of Clusters K

(a) *February 28, 2014*

(b) *April 1, 2014*

(c) *May 19, 2014*

(d) *June 20, 2014*

(e) *July 22, 2014*

(f) *August 7, 2014*
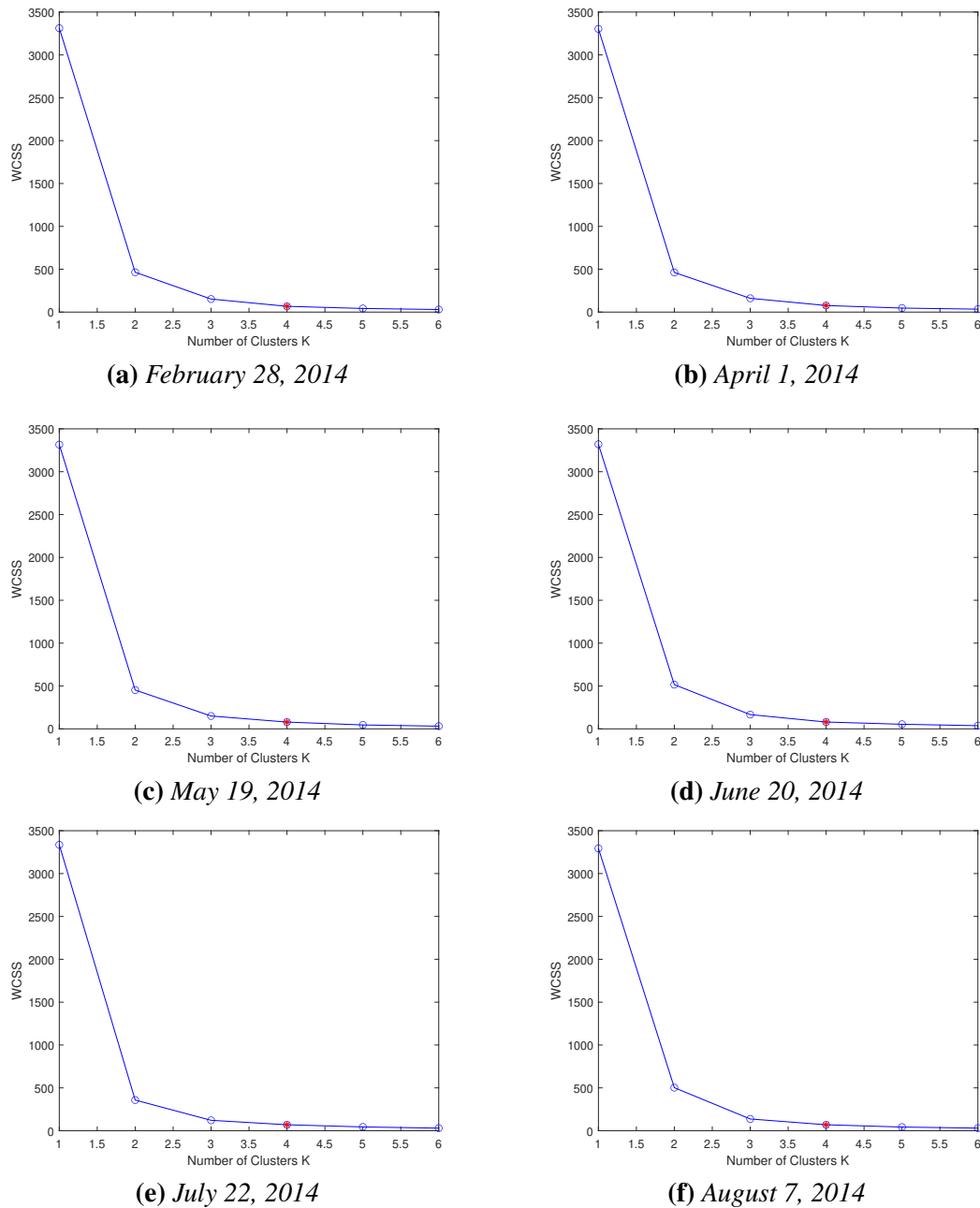
The separating measures how well the clusters are separated. The silhouette S_i for each observation i, is defined as follows [60]:

1. The cohesion a_i for each observation i, measures the average distance a_i between

57

observation i and all other points of the same cluster C_i (similarity).

$$a_i = \frac{1}{C_i - 1} \sum_{j \in C_i, i \neq j} d(i, j) \tag{23}$$

Where, C_i is the number of observation in cluster i and $d(i, j)$ is the distance between observations i, j in the same cluster C_i. The smaller the value of a_i the better is the cluster assignment of observation i.

2. The separation b_i for each observation i, measures the average distance of i to all other observations in other clusters C (dissimilarity).

$$b_i = \min_{k \neq i} \frac{1}{C_k} \sum_{j \in C_k} d(i, j) \tag{24}$$

where, K is the cluster number and $(C_k \neq C_i)$. Therefor, for each observation $i \in C_i$ we find the minimum average distance to all points j in cluster C_k.

3. The silhouette values S_i for each observation i is defined by:

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{25}$$

The silhouette values ranges from $-1 <= S_i <= 1$. A higher values indicate that observation is assigned to the right cluster while negative values indicate that the observation is assigned to the wrong cluster. A small silhouette score around zero, indicates that the observation lies between two clusters. The final silhouette coefficient is the mean of s_i over all observations of the entire dataset for a specific number of clusters K.

7.2.1 Silhouette Analysis for Conventional K-means and VAE K-means clustering

Figure 19 shows the silhouette analysis for both the K-means - left column - and VAE K-means - right column - for the same images in Figure 14 row 2 and 3 respectively.

58

Again, the silhouette analysis clearly show that the VAE features enhance the silhouette co-efficient (mean of silhouette for all data point in one observation) by 10-16%. The negative silhouette values show some data points are incorrectly assigned to that cluster. Tracing some of these points were found to be on the edge between two clusters, for example their values would lie between cluster one (no risk) and cluster two (lower risk) and it will not effect our analysis if it they are assigned to either one of them.

7.3 Summary of the Internal Clustering Validation

Table 3 summarizes the internal clustering validation and determines the optimal number of clusters. Elbow analysis results are presented in columns (2-4). The optimal cluster number is $K = 4$ for K-means and VAE K-means. Columns (5-8) show the results of K-means clustering using silhouette analysis. The results show that the optimal cluster number is $K = 2$ for most observations. However, the difference between silhouette co-efficients of $K = 2$ and $K = 4$ is minimal as shown in the difference column in Table 3 and Figure 19. The silhouette coefficients for VAE K-means is shown in column (9-12). The optimal number of clusters is found to be $K = 2$ except for 2 observations. The difference between silhouette coefficients of $K = 2$ and $K = 4$ is again very small (1%-3%). The results also clearly show that the VAE K-means for the salient features of the encoder output of VAE can enhance the results of the conventional K-means results. It shows that the silhouette coefficients for $K = 2$ is improved by 7-11% while for $K = 4$ by 10-16%.

Referring to the basic definition of K-means in Section 1.3, the algorithm learns patterns from the data given a preset number of clusters (k). The number of clusters can be inferred from the knowledge of what the data represents and the problem at hand. Furthermore, the clustering validation methods, elbow and silhouette, give an indication of the optimal number of clusters for the dataset, which can then be used to further strengthen our confidence in the selected number of cluster (k).

59

The elbow method is more of a decision rule, whereas the Silhouette is a metric used for validation during clustering. As a result, it may be used with the elbow method. The elbow method and the silhouette method are not interchangeable when it comes to determining the optimum K. Rather, they are methods to be utilized in tandem to make a more confident conclusion about the clustering results.

Therefore the assumption of $K = 4$, which is inferred from the knowledge and understanding of the data as well as the problem at hand, is supported by the internal clustering validation methods demonstrated above, where the results of both methods are withing range of our assumption of $K = 4$. Dividing the landfill to four clusters, defines the gradual change in temperature over time. This is crucial to assess the landfill area to consider and assign the appropriate actions to be taken given it's status.

60

Figure 19

The Silhouette Analysis for Observation Form February to August 2014
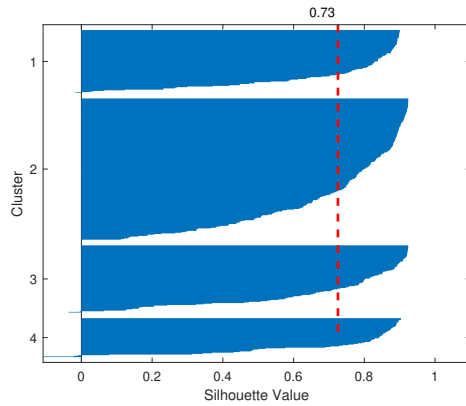
(a) *February 28, 2014*
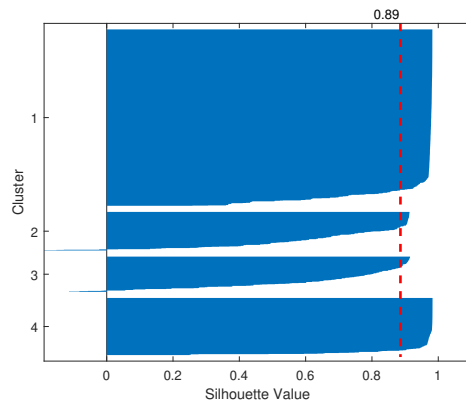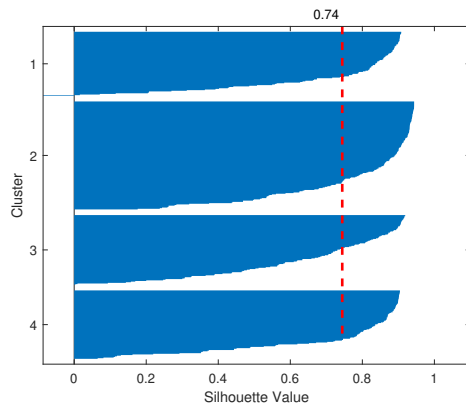
(b) *February 28, 2014*

(c) *April 1, 2014*

(d) *April 1, 2014*

(e) *May 19, 2014*

(f) *May 19, 2014*

61
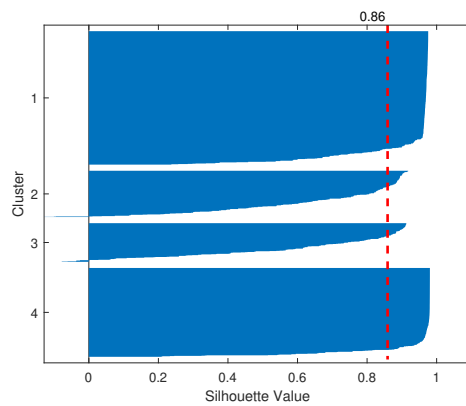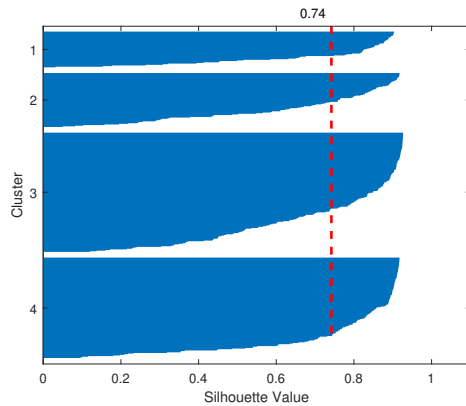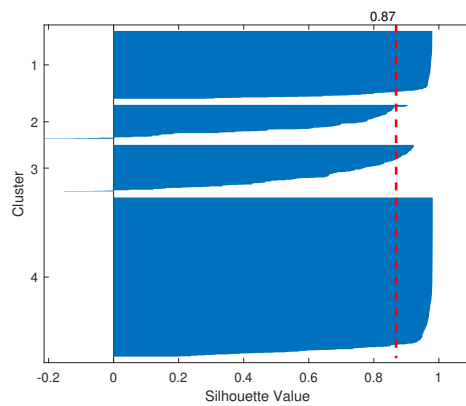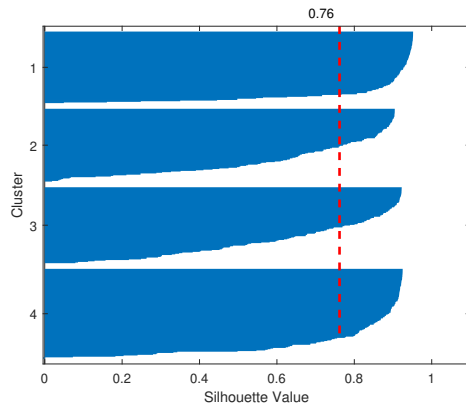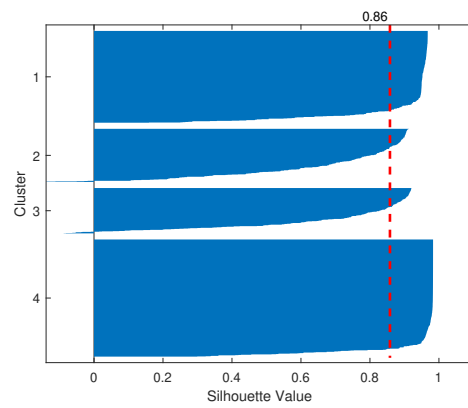
(g) *June 20, 2014*

(h) *June 20, 2014*
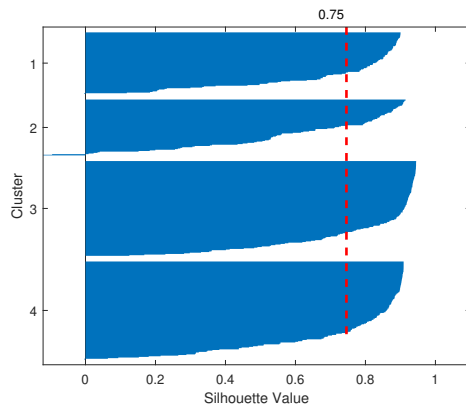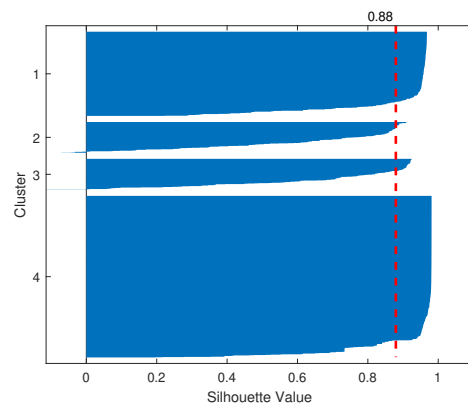
(i) *July 22, 2014*

(j) *August 7, 2014*
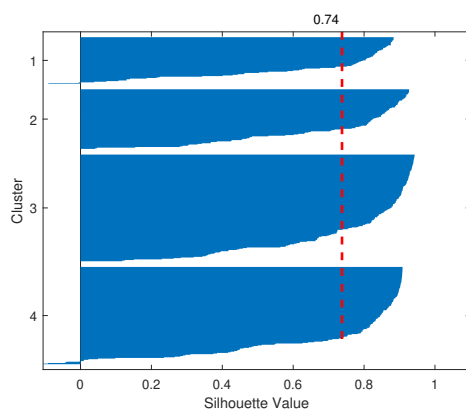
(k) *July 22, 2014*
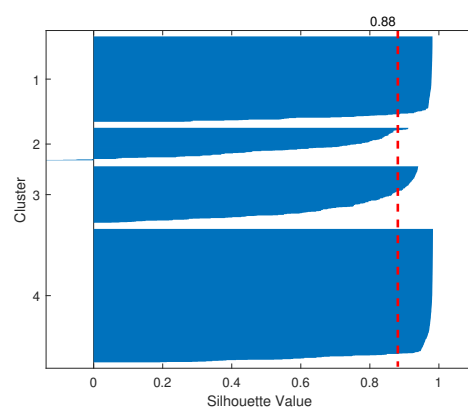
(l) *August 7, 2014*

Figures (a, c, e, g, i and k) are for temperature data as in Figure 14 row 2. Figures(b, d, f, h, j and l) are for VAE extracted features as in Figure 14 row 3. The red lines indicate the silhouette coefficient for the clustered data.

62

Table 3

Summary of the Internal Clustering Validation

Image Date	Elbow			Silhouette				VAE K-mean			
	K-means	VAE K-means 90%	VAE K-means 95%	K-means							
				Optimal K ^{a}	k=2	k=4	Diff.	Optimal K ^{b}	k=2	k=4	Diff.
February 28, 2014	4	3	4	2	0.79	0.73	-0.07	2	0.90	0.89	-0.01
April 1, 2014	4	3	4	2	0.77	0.74	-0.02	2	0.88	0.86	-0.02
May 19, 2014	4	3	4	2	0.82	0.74	-0.08	2	0.89	0.87	-0.02
June 20, 2014	3	3	4	3	0.76	0.76	0.00	3	0.86	0.86	0.00
July 22, 2014	4	3	4	2	0.83	0.75	-0.08	2	0.91	0.88	-0.03
August 7, 2014	4	3	4	2	0.80	0.74	-0.06	3	0.87	0.88	0.01

^{a}Obtained by applying silhouette analysis from $k = 1 : 6$ to temperature data.
^{b}Obtained by applying silhouette analysis from $k = 1 : 6$ to VAE features.

63

Chapter 8

Simplified Framework for Quantifying Landfill Health State

In chapter 5 and chapter 6, we have shown different clustering algorithms; Heat Index (HI) based on the standard deviation from the mean, classical unsupervised K-means and VAE K-means using unsupervised deep learning model. All these algorithms resulted in the landfill being clustered to four different areas denoted as no risk, lower risk, moderate risk and higher risk for each individual LST observation. Regardless of the clustering method used, they can all be used to quantify the health state of the landfill. The quantification of the health state of the landfill not only allows one to evaluate the current state of the landfill but also to shed light on past events and to predict where the next heat elevation or possible fire will occur. The block diagram in Figure 20, describes the overall algorithm implementation and Algorithm 3 gives more details.

Figure 20

Block Diagram of Algorithm 3.(Left Block) Extracting Health State Indices;(Right Block) Plotting Indices
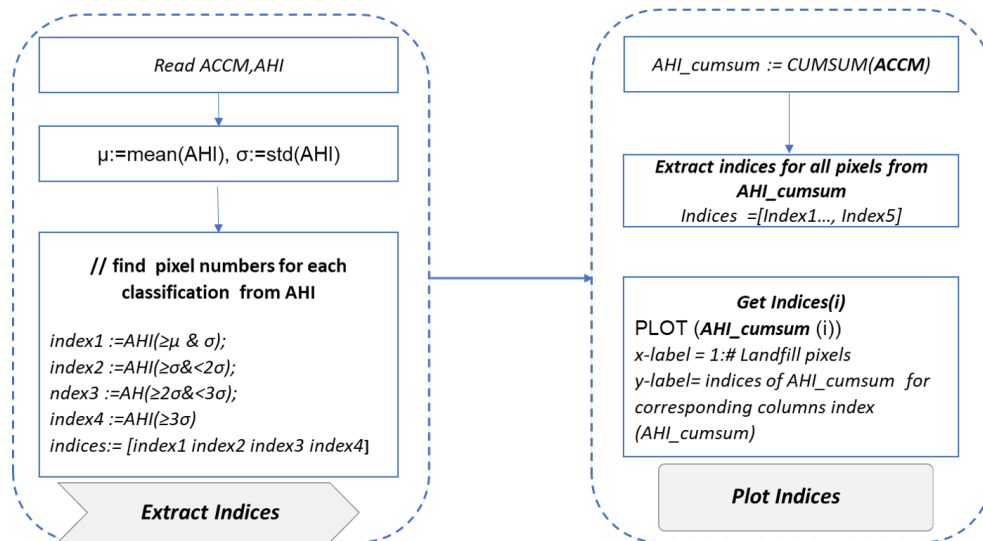
The block diagram of Algorithm 3 has two main parts, Part one, shown on the left block, uses the matrix *ACCM* generated by HI as described in chapter 5 to assign a new indices that define the four categories of landfill state (NO Risk, Lower Risk, Moderate Risk and Higher Risk) at any given date. Part two, the right block of in Figure 20 is used to plot these indices.

The health state quantifying process described in Algorithm 3, explores the historical behavior of each pixel in all LST observations by tracing the temperature profile for any given pixel during the 20 years of study and plotting each index value and its corresponding pixel's accumulated index values. In our previous work [14], this analysis was able to show the SSE and the overall heat elevation in Bridgeton, Missouri Landfill in 2010.

ACCM is a matrix ($n \times m$), where n is the number of pixels in the landfill and m is the number of observations that stores all HI indices for one observation row wise. *AHI* is a matrix ($l \times w$) of the landfill size, where l and w are the length and width of the landfill respectively. It is obtained by summing LST observations as they are read. Averaging and re-indexing the content of *AHI* at any give date shows the average clustering for all the previous observation until that date.

Algorithm 3: Indices to quantify landfill Health state

Input: *AHI*, *ACCM*　　//From Algorithm 2

Output: Plot Landfill Health State Indices

1　$\mu := mean(AHI)$, $\sigma := std(AHI)$

　　/* Get pixel numbers corresponding to index　　　　　　　　　　　　　　　*/

2　$index1 := AHI(\geq \mu \,\&\, \sigma)$　　　　　　　// pixels numbers with index1 (No Risk)

3　$index2 := AHI(\geq \sigma \,\&\, < 2\sigma)$　　　// pixels numbers with index2 (Lower Risk)

4　$index3 := AHI(\geq 2\sigma \,\&\, < 3\sigma)$　// pixels numbers with index3 (Moderate Risk)

5　$index4 := AHI(\geq 3\sigma)$　　　　　　// pixels numbers with index4 (Higher Risk)

6　$indices := [index1 \; index2 \; index3 \; index4]$

7

$$ACCM_{cumsum} = CUMSUM(ACCM) \tag{26}$$

　　/* Where $ACCM_{cumsum}$, is the column wise accumulated sum of $ACCM$ and

　　$CUMSUM$ is the cumulative sum operation.　　　　　　　　　　　　　　*/

8　**Function** PLOT($ACCM_{cumsum}$, *indices*)**:**

　　　/* Find the equivalent pixel numbers from columns I_j　　　　　　　*/

9　　　$j := 1$

10　　**while** $< length(indices)$ **do**

11　　　　$I_j := indices(j)$　　　　　　　　　　// pixel numbers given by $index(j)$

12　　　　PLOT $(ACCM_{cumsum}(I_j))$　// Plot columns from $ACCM_{cumsum}$ specified by

　　　　$I_j)$

13　　　　$j := J + 1$

14　　　　**return** (Plots as in Figure 21 and Figure 22)

15

Figure 21

Accumulated Heat Index for Deerfield Landfill by the End of Study Period on 25-11-2019

(a) *Index = 1 (43.93%)*

(b) *Index = 2 (35.63%)*

(c) *Index = 3 (20.45%)*
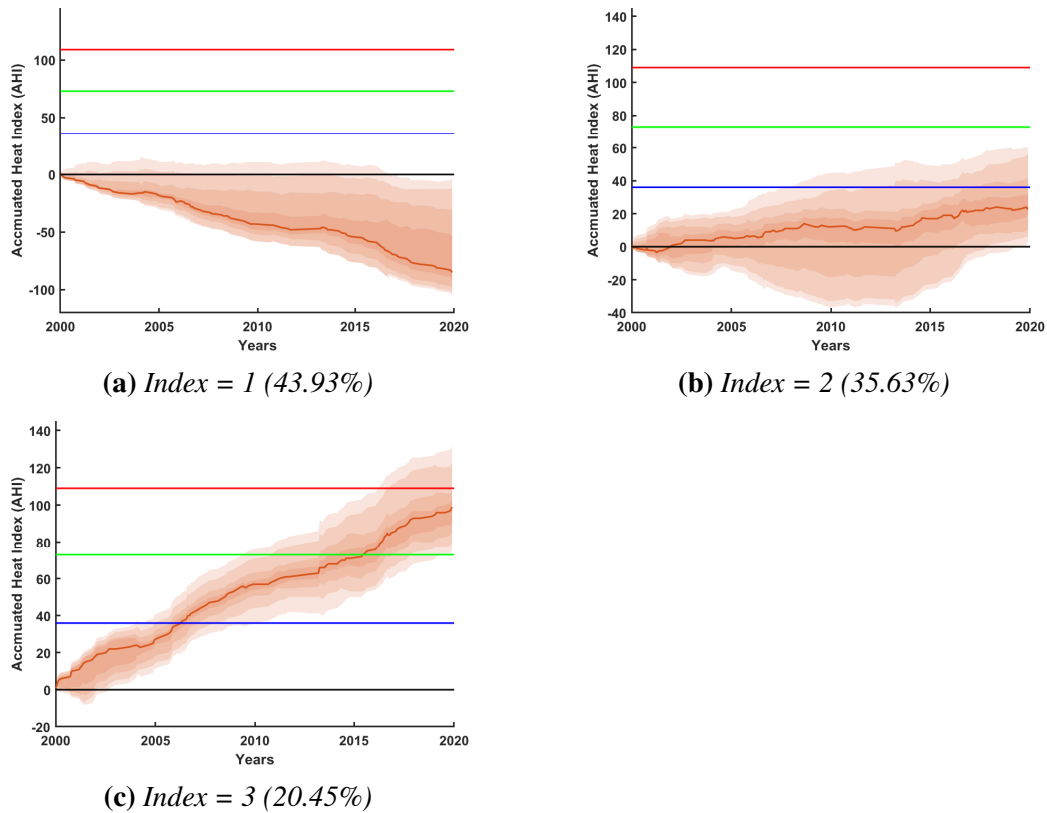
The result of this analysis is shown in Figure 21 (a-c) for Deerfield landfill and Figure 22 (d-e) for South Harrison landfill. It shows the percentages (%) of the total landfill pixels of the indices (1-4) corresponding to the states (no risk, lower risk, moderate risk and higher risk) on 25-11-2019. The vertical axis shows the accumulated heat index, where negative numbers indicate the number of times a pixel has recorded a temperature below the mean and the positive numbers indicate the accumulated index until a given date. A closer look at Figure 21 (a-c) shows that Deerfield landfill kept a healthy thermal state for 20 years. However, pixels having index = 3 (moderate risk) comprise 20.5% of total pixels in the landfill, which could be considered as a warning. Index = 1 (44%) of total pixels in the landfill shows temperature below the mean most of the time. Index = 2 (35.5%) of the pixels in the landfill. Although, some pixels in this area show low temperature (below the

67

Figure 22

Accumulated Heat Index for South Harrison Landfill by the End of Study Period on 25-11-2019

(a) *Index = 1 (51.46%)*

(b) *Index=2 (30%)*

(c) *Index = 3 (16.5%)*

(d) *Index = 4 (2.0%)*
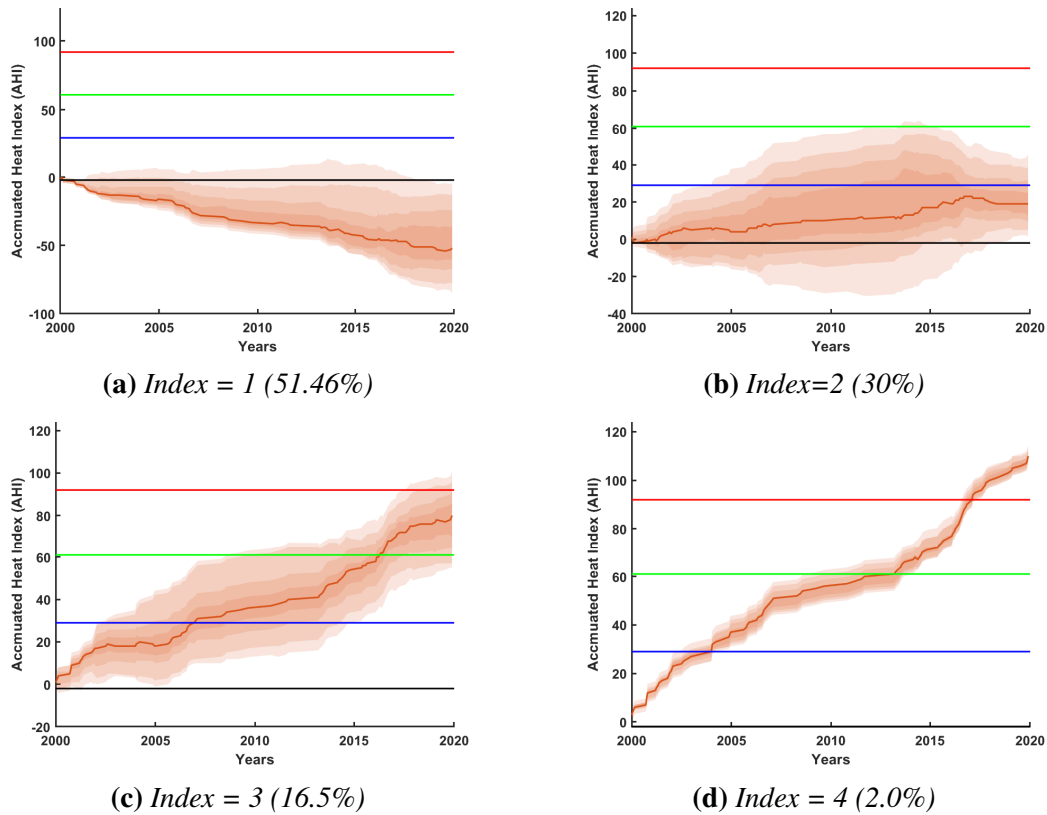
mean) that starts to increase from 2013, other pixels have already increasing temperature behavior and should be monitored. Indices 1, 2 and 3 are the same indices in Figure 9e represented by the colors blue, green and orange, respectively. It should be noted that Deerfield landfill has no index = 4, which means that it is still maintaining a healthy state and no fires have been reported as far as we know. The same analysis is applied to South Harrison Landfill in Figure 22 (d-e), where a few pixels denoted by index = 4 (2%) of the total pixels of the landfill show increasing temperatures since the beginning of the study. Even though this landfill is closed since 2012, this area of the landfill still maintains high temperature, which may continue for the following years. The general assessment for this landfill reports a healthy state during its operation period. 51.5% maintaining -no risk- denoted by index = 1. 30% of the landfill has -lower risk- and only 16.5% of the landfill

show -moderate risk- but still below any warning levels. There is only (2%) of the landfill classified as higher risk that need to be monitored.

In Figure 21 we showed Deerfield landfill state on November 11, 2019 the last observation, which is row 1 in Table 4 that concludes all the previous observations since the year 2000. Figure 23 shows the Deerfield landfill health state in the past 20 years (2000-2020). This shows the changes in percentages of landfill area of different risk levels. Through the changes in the indices that indicate the risk levels, we can predict the thermal behavior of this landfill as well as estimate its thermal behavioral patterns in past years even if there's no data available to study. Deerfield landfill has maintained a relatively healthy state throughout the years of the study keeping its risk level from no risk to moderate risk. However, there were times when the landfill showed higher temperatures with Index = 4 corresponding to a higher risk level in a couple of years. Through appropriate maintenance of the field and environmental changes the temperatures were brought down to lower risk levels.

Figure 23

Deerfield Landfill Thermal State Patterns From Year 2000-2019
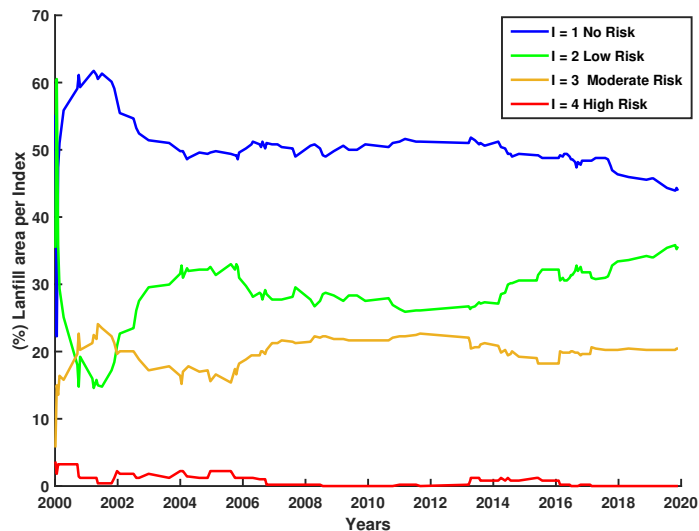
8.1 The Interpretation of Heat Indices and Quantification of Landfill Healthiness

Heat indices, as explained above, are a quantification of the behavior of each pixel in the landfill. They are determined by tracing the temperature profile for any given pixel during the 20 years of study and plotting each index and its corresponding pixels as shown in Algorithms 1, 2, Figure 21 and Figure 22. Consequently, it is a measure of the health of the landfill at any given time. To finalize this analysis, we show the summary of each index percentage in Table 4 for the landfills listed in Table 1. Table 4 evaluates the health of the landfills in South New Jersey, USA, by the end of study period (25/11/2019).

Generally, I=3 and 4 tend to have an increasing behavior during the whole study period as shown in Figures 12 and 13. The more the percentage of I = 3 and I = 4 the less healthy is the landfill. Increasing (%) of I=3 could be an indication that a smoldering event is about to occur in that area of the landfill, especially if these indices turn to I = 4 with continuous increase of temperature. Areas with persistent I = 4 is an indication of ongoing SEE. The more percentage of I=1, the healthier the landfill and it is not a matter of concern as they are always below or around the mean. The color codes in Table 4 are the same colors used in K-means and Heat Index. The same analysis is applied to Bridgeton Sanitary Landfill, MO, USA where the indices for the Bridgeton, MO landfill were found to be, I = 4 (25%), I = 3 (10%), and the rest of the indices I = 1 and I = 2 are 65%. The subsurface fire was first identified in 2010 where the SSE reported can be inferred from all the indices reaching their maximum levels and continues to burn until today. This landfill was under scrutiny for a long time due to the continuous fires, smoke and odors that erupted from it. Also in [14], we could identify hotspots reported in 2014 and SSE reported in 2012, which explains the continuous heat elevation over the entire landfill. This technique can effectively detect most hotspots and the results have been verified by a consultant report [61, 62].

Table 4

Summary of Percentage per Pixel per Index for Landfills Listed in Table 1 by the End of Study Period (25/11/2019) Using the Method Described by Algorithm 2 and 3

Landfill Name	I=1	I=2	I=3	I=4
	No Risk	Lower Risk	Moderate Risk	Higher Risk
Deerfield Township	43.93%	35.63%	20.45%	0%
South Harrison Township	51.46%	30.05%	16.49%	2.00%
Mannington Township	59.74%	16.58%	23.68%	0%
Egg Harbor Township	37.50%	47.64%	13.21%	1.65%
Woodbine Borough	56.27%	23.57%	19.01%	1.14%
Carney's Point Township	46.05%	34.87%	17.11%	1.97%
Vineland City	54.98%	26.54%	11.85%	6.64%
Millville City	34.09%	51.14%	10.23%	4.55%
Commercial Township	47.04%	33.99%	15.81%	3.16%

Chapter 9

Discussion and Conclusions

This study proposed a deep learning technique derived from satellite observations to detect anomalies in surface temperature in active as well as closed landfills that may lead to landfill fires. It can be used to address the problem of locating hotspots by monitoring the thermal signature of these waste sites. In this work, a noninvasive and cost-effective method is proposed for monitoring temperature changes through the collection and analysis of satellite imagery. This overcomes the lack of any ground truth data from individual landfills, and no expenditure of any devices nor manpower, and without relying on any method that is otherwise costly or time consuming to enable the timely detection of subsurface smoldering events. To reach this goal, temperature data contained in the Landsat satellite images were converted into a more workable format and then analyzed.

To address the problem of the absence of onsite observations, one of the main goals of this study was to demonstrate applicability and advantages of remote sensing data coupled with machine learning techniques necessary to identify landfill thermal states that can lead to fire events. On one hand, remote sensing can be used to locate hotspots by monitoring the thermal signature of these landfills. On the other hand, the machine learning algorithms will address the problem of the missing ground truth data (labeled data) by applying unsupervised machine learning methods to detect the thermal states of the landfills and to detect anomalies. As we have described above, the unsupervised machine learning algorithms are able to detect the hidden patterns and cluster them without the need for human intervention [1]. Unsupervised learning pass large volumes of unstructured data to algorithms or neural networks, enabling them to learn, infer and find relations in the given data.

As presented in the results section, the location of hotspots at nine landfills in South New Jersey, were successfully detected and monitored using different clustering algorithms; Heat Index (HI) based on the standard deviation from the mean, classical unsupervised K-means and VAE K-means using unsupervised deep learning model. We used the internal clustering validation tools such as Silhouette and Elbow to quantitatively show the accuracy of our clustering results which proved VAE K-means clustering method to be superior to the classical unsupervised K-means clustering method. The Silhouette coefficient of VAE K-means of K=2 and K=4 had improved 7-11% and 10-11% respectively as compared to the classical K-means. While the Elbow method showed the WCSS range to be 500-2000 for the classical K-means and 50-200 for the VAE K-means, indicating that VAE K-means clusters are more coherent to their centroid. All these algorithms were able to cluster the landfill into four different areas denoted as no risk, lower risk, moderate risk and higher risk for each individual LST observation. Regardless of the method, We developed a simple framework to quantify the health state of the landfill as shown in Figure 22 and Table 4. The quantification of the health state of the landfill not only allows to evaluate the current state of the landfill but also to shed light on past events and to predict where the next heat elevation or possible fire will occur.

The use of satellite remote sensing techniques for the detection of possible fires in landfills has practical significance when there is no on-site landfill data available or in the detection of illegal waste dumps. The 30-m spatial resolution of the thermal band can detect most of the substantial hotspots as these usually last for months and their generated heat propagates both vertically and horizontally for distances that are detectable by satellite infrared sensors. However, we enhanced the resolution of generated LST observations to approximately 2-m using appropriate 2D interpolation.

Hotspots take weeks, months or even years to develop. For this reason, the length of time between revisits of the satellite (every 16 days) and missing observations due to cloud coverage are not an issue. However, to minimize the limitations of satellite availability, unmanned aerial vehicles (UAV's) such as drones can effectively be deployed, and the same algorithms developed herein can still be utilized.

Future work will use the results generated by this study to provide data input for a monitoring system that can be used to issue warnings regarding potential landfill fires and to identify anomalous thermal patterns and changes of any landfill. The results also provide new datasets that can be used for further investigation using deep learning approaches. Our future work will incorporate more advanced deep learning techniques to detect anomalies directly from thermal bands such as B10 and B11 in Landsat 8. This will save ample time in calculating LST and the thermal bands can be directly fed to VAE to detect anomalies. These results can be used in recurrent neural network (RNN) and Long short-term memory (LSTM) networks to predict the thermal state of the landfill.

Thermal remote sensing is an effective tool for monitoring the internal activities of landfills and provides a reliable method for predicting fire outbreaks and preventing possible environmental disasters. Given the availability of public data from the USGS Explorer satellite images database, the proposed method can be applied to any landfill in USA territory to predict subsurface thermal events.

References

[1] J. Delua, *Supervised vs. unsupervised learning: What's the difference?* IBM Education, blog, Mar. 2021. [Online]. Available: https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning (visited on 05/31/2021).

[2] P. Dangeti, *Statistics for Machine Learning*. Packt Publishing Ltd, 2017, ISBN: 978-1-78829-575-8.

[3] S. Mishra, *Unsupervised learning and data clustering*, Towards data science, blog, May 2017. [Online]. Available: https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a (visited on 05/31/2021).

[4] T. H. Robert, Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, ser. Springer Series in Statistics. Springer, 2008, ISBN: 978-0387848570.

[5] R. Sharma, *What is clustering and different types of clustering methods*, Upgard, blog, Dec. 2020. [Online]. Available: https://www.upgrad.com/blog/clustering-and-types-of-clustering-methods (visited on 05/31/2021).

[6] S. Kaushik, *An introduction to clustering and different methods of clustering*, Analytic Vidha, blog, Nov. 2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/ (visited on 05/31/2021).

[7] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009. [Online]. Available: https://doi.org/10.1145/1541880.1541882 (visited on 05/31/2021).

[8] K. Ring Burbeck, *Adaptive real-time anomaly detection for safeguarding critical networks*, Report code: LiU-Tek-Lic-2006:12., 2006.

[9] H. You, Z. Ma, Y. Tang, Y. Wang, J. Yan, M. Ni, K. Cen, and Q. Huang, "Comparison of ann (mlp), anfis, svm, and rf models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators," *Waste Management*, vol. 68, pp. 186–197, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0956053X17301903.

[10] J. L. Hanson, W.-L. Liu, and N. Yesiller, "Analytical and numerical methodology for modeling temperatures in landfills," in *GeoCongress 2008*, pp. 24–31. DOI: 10.1061/40970(309)3. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/40970%28309%293.

[11] H. Alfergani, R. Nazari, and N. Bouaynaya, "Application of remote sensing and deep learning in detecting internal temperature anomalies in landfills," unpublished.

[12] J. Kret, L. Dalidowitz Dame, N. Tutlam, R. W. DeClue, S. Schmidt, K. Donaldson, R. Lewis, S. E. Rigdon, S. Davis, A. Zelicoff, C. King, Y. Wang, S. Patrick, and F. Khan, "A respiratory health survey of a subsurface smoldering landfill," *Environmental Research*, vol. 166, pp. 427–436, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0013935118302779.

[13] K. Youmaran, "We didn't start the fire: The current outlook of the bridgeton landfill and its implication for missourians," *ENVTL. & SUSTAINABILITY*, vol. 22, pp. 365–382, 2016. [Online]. Available: https://scholarship.law.missouri.edu/cgi/viewcontent.cgi?article=1477&context=jesl.

[14] R. Nazari, H. Alfergani, F. Haas, M. E. Karimi, M. G. R. Fahad, S. Sabrin, J. Everett, N. Bouaynaya, and R. W. Peters, "Application of satellite remote sensing in monitoring elevated internal temperatures of landfills," *Applied Sciences*, vol. 10, no. 19, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/19/6801.

[15]   Federal Emergency Management Agency, "Landfill fires. their magnitude, characteristics, and mitigation," Agency for Toxic Substances and Disease Registry, Arlington,, Virginia, USA, Tech. Rep. FA–225, 2002. [Online]. Available: https://www.sustainable-design.ie/fire/FEMA-LandfillFires.pdf (visited on 06/04/2020).

[16] J. T. Powell, T. G. Townsend, and J. B. Zimmerman, "Estimates of solid waste disposal rates and reduction targets for landfill gas emissions," *Nature Climate Change*, vol. 6, no. 2, pp. 162–165, Feb. 2016. DOI: 10.1038/nclimate2804.

[17] D. Gasbarra, P. Toscano, D. Famulari, S. Finardi, P. Di Tommasi, A. Zaldei, P. Carlucci, E. Magliulo, and B. Gioli, "Locating and quantifying multiple landfills methane emissions using aircraft data," *Environmental Pollution*, vol. 254, p. 112 987, 2019, ISSN: 0269-7491. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749119309455.

[18] A. Kerric, J. Okeme, L. Jantunen, J.-F. Giroux, M. L. Diamond, and J. Verreault, "Spatial and temporal variations of halogenated flame retardants and organophosphate esters in landfill air: Potential linkages with gull exposure," *Environmental Pollution*, vol. 271, p. 116 396, 2021, ISSN: 0269-7491. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0269749120370858.

[19] V. R. Propp, A. O. De Silva, C. Spencer, S. J. Brown, S. D. Catingan, J. E. Smith, and J. W. Roy, "Organic contaminants of emerging concern in leachate of historic municipal landfills," *Environmental Pollution*, vol. 276, p. 116 474, 2021, ISSN: 0269-7491. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026974912100052X.

[20] M. Brune, H. G. Ramke, H. J. Collins, and H. H. Hanert, "Incrustation processes in drainage systems of sanitary landfills," in *3rd Int. Landfill Symp., Cagliari)*, 1991, pp. 999–1035.

[21] P. Döll, "Desiccation of mineral liners below landfills with heat generation," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 123, no. 11, pp. 1001–1009, 1997. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%291090-0241%281997%29123%3A11%281001%29.

[22] M. El-Fadel, A. Findikakis, and J. Leckie, "Numerical modelling of generation and transport of gas and heat in sanitary landfills ii. model application," *Waste Management & Research*, vol. 14, no. 6, pp. 537–551, 1996. DOI: 10.1177/0734242X9601400603. [Online]. Available: https://doi.org/10.1177/0734242X9601400603.

[23] S. Sabrin, R. Nazari, M. Karimi, G. R. Fahad, J. Everett, and R. Peters, "Development of a conceptual framework for risk assessment of elevated internal temperatures in landfills," *Science of The Total Environment*, vol. 782, p. 146 831, Mar. 2021. DOI: 10.1016/j.scitotenv.2021.146831.

[24] J. W. Atwater, "Solid waste landfill engineering and design," *Canadian Journal of Civil Engineering*, vol. 23, no. 3, pp. 647–648, 1996. DOI: 10.1139/l96-879. [Online]. Available: https://doi.org/10.1139/l96-879.

[25] S. J. Pirt, "Aerobic and anaerobic microbial digestion in waste reclamation," *J. APPL. CHEM. BIOTECHNOL*, 1978.

[26] J. Southen and R. K. Rowe, "Modelling of thermally induced desiccation of geosynthetic clay liners," *Geotextiles and Geomembranes*, vol. 23, no. 5, pp. 425–442, 2005, ISSN: 0266-1144. DOI: https://doi.org/10.1016/j.geotexmem.2005.01.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0266114405000245.

[27] H. Yoshida and R. K. Rowe, "Consideration of landfill liner temperature," in *Proceedings Sardinia 2003, Ninth International Waste Management and Landfill Symposium*, 2003.

[28] N. H. Jafari, T. D. Stark, and T. Thalhamer, "Progression of elevated temperatures in municipal solid waste landfills," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 143, no. 8, p. 05 017 004, 2017.

[29] D. Hall, D. Drury, R. Keeble, A. Morgans, and R. Wyles, "Review and investigation of deep-seated fires within landfill sites," *Environment Agency: Bristol, UK*, 2007.

[30] N. Yeşiller, J. L. Hanson, and W.-L. Liu, "Heat generation in municipal solid waste landfills," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 131, no. 11, pp. 1330–1344, 2005. DOI: 10.1061/(ASCE)1090-0241(2005)131:11(1330).

[31] A. Y. Kwarteng and A. Al-Enezi, "Assessment of kuwait's al-qurain landfill using remotely sensed data," *Journal of Environmental Science and Health, Part A*, vol. 39, no. 2, pp. 351–364, 2004. DOI: 10.1081/ESE-120027527. [Online]. Available: https://doi.org/10.1081/ESE-120027527.

[32] A. Shaker and W. Y. Yan, "Trail road landfill site monitoring using multitemporal landsat satellite data," in *Canadian Geomatics Conference 2010 and ISPRS , (COM I Symposium)*, 2011, pp. 95–106.

[33] H. Alfergani, N. Bouaynaya, and R. Nazari, "Spatio-temporal statistical sequential analysis for temperature change detection in satellite imagery," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 2917–2920. DOI: 10.1109/IGARSS39084.2020.9323164.

[34] D. Lelescu and D. Schonfeld, "Statistical sequential analysis for real-time video scene change detection on compressed multimedia bitstream," *IEEE Transactions on Multimedia*, vol. 5, no. 1, pp. 106–117, 2003. DOI: 10.1109/TMM.2003.808819.

[35] J. Zielinski and N. Bouaynaya, "Statistical sequential analysis for detection of microcalcifications in digital mammograms," in *International Conference on Signal Processing and Communications (SPCOM)*, Jul. 2010, pp. 1–5.

[36] H. Hong-Yuan, J. Xiaoguang, S. Xianfeng, N. Zhuoya, and G. Caixia, "Detection of coal fire dynamics and propagation direction from multi-temporal nighttime landsat swir and tir data: A case study on the rujigou coalfield, northwest (nw) china," *Remote Sensing*, pp. 1234–1259, Jan. 2014.

[37] H. Hong-Yuan, J. Xiaoguang, N. Zhuoya, G. Caixia, Z. Enyu, Y. Zhang, L. Yi, Z. Huili, Z. Shiyue, J. Xiaoguang, S. Xianfeng, Z. Ping, and C. Tiejun, "A study of coal fire propagation with remotely sensed thermal infrared data. remote sensing," *Remote Sensing*, vol. 7, pp. 3088–3113, Mar. 2015.

[38] C. Kuenzer, J. Zhang, J. Li, S. Voigt, H. Mehl, and W. Wagner, "Detecting unknown coal fires: Synergy of automated coal fire risk area delineation and improved thermal anomaly extraction," *International Journal of Remote Sensing*, vol. 28, no. 20, pp. 4561–4585, Sep. 2007.

[39] K. Claudia, Z. Jianzhong, L. Jing, G. Huadong, and D. Stefan, "Thermal infrared remote sensing of surface and underground coal fires," *Thermal Infrared Remote Sensing. Remote Sensing and Digital Image Processing.*, vol. 17, pp. 429–451, 2013.

[40] C. Varun and V. Ranga, "A gaussian process based online change detection algorithm for monitoring periodic time series," in *Proceedings of the 11th SIAM International Conference on Data Mining, (SIAM)*, Apr. 2011, pp. 95–106.

[41] S. K. Yarlagadda, D. Güera, P. Bestagini, F. M. Zhu, S. Tubaro, and E. J. Delp, *Satellite image forgery detection and localization using gan and one-class classifier*, 2018. arXiv: 1802.04881 [cs.CV].

[42] R. Yao, C. Liu, L. Zhang, and P. Peng, "Unsupervised anomaly detection using variational auto-encoder based feature extraction," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2019, pp. 1–7. DOI: 10.1109/ICPHM.2019.8819434.

[43] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, *Adversarially learned one-class classifier for novelty detection*, 2018. arXiv: 1802.09088 [cs.CV].

[44] Survey, U.S.G. and Department of the Interior (US), "Landsat 8 (L8) data users handbook (V2.0)," U.S. Geological Survey Landsat Missions, Tech. Rep. LSDS-1574, 2016.

[45] Q. Weng, D. Lu, and J. Schubring, "Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies," *Remote Sensing of Environment*, vol. 89, no. 4, pp. 467–483, 2004, ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2003.11.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425703003390.

[46] A. A. V. D. GRIEND and M. OWE, "On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces," *International Journal of Remote Sensing*, vol. 14, no. 6, pp. 1119–1131, 1993. [Online]. Available: https://doi.org/10.1080/01431169308904400.

[47] J. Zhang, Y. Wang, and Y. Li, "A C++ program for retrieving land surface temperature from the data of landsat TM/ETM+ band6," *Comput. Geosci.*, vol. 32, no. 10, pp. 1796–1805, Dec. 2006, ISSN: 0098-3004. DOI: 10.1016/j.cageo.2006.05.001. [Online]. Available: https://doi.org/10.1016/j.cageo.2006.05.001.

[48] J. A. Sobrino, J. C. Jiménez-Muñoz, and L. Paolini, "Land surface temperature retrieval from landsat tm 5," *Remote Sensing of Environment*, vol. 90, no. 4, pp. 434–440, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425704000574.

[49] Z.-L. Li, H. Wu, N. Wang, S. Qiu, J. A. Sobrino, Z. Wan, B.-H. Tang, and G. Yan, "Land surface emissivity retrieval from satellite data," *International Journal of Remote Sensing*, vol. 34, no. 9-10, pp. 3084–3127, 2013. DOI: 10.1080/01431161.2012.716540. [Online]. Available: https://doi.org/10.1080/01431161.2012.716540.

[50] E. S. R. Institute, *Change detection*, Environmental Systems Research Institut, website. [Online]. Available: https://support.esri.com/en/other-resources/gis-dictionary/term/55a5d9c1-1138-4da5-8152-5060b176b133 (visited on 05/31/2021).

[51] S. Aminikhanghahi and D. Cook, "A survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, May 2017. DOI: 10.1007/s10115-016-0987-z.

[52] A. B. Downey, *A novel changepoint detection algorithm*, 2008. arXiv: 0812.1237 [stat.AP].

[53] B. Namoano, A. Starr, C. Emmanouilidis, and R. C. Cristobal, "Online change detection techniques in time series: An overview," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2019, pp. 1–10. DOI: 10.1109/ICPHM.2019.8819394.

[54] Z. Govindarajulu, "Chapter 1 - introduction and certain double sampling procedures," in *Sequential Statistical Procedures*, ser. Probability and Mathematical Statistics: A Series of Monographs and Textbooks, Z. Govindarajulu, Ed., Academic Press, 1975, pp. 1–13. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780122942501500071.

[55] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc-Verlag, 1993.

[56] Agency for Toxic Substances and Disease Registry, "Landfill gas primer: An overview for environmental health professionals," Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, Tech. Rep., 2001. [Online]. Available: https://www.atsdr.cdc.gov/HAC/landfill/html/toc.html (visited on 06/04/2020).

[57] S. Odaibo, *Tutorial: Deriving the standard variational autoencoder (vae) loss function*, 2019. arXiv: 1907.08956 [cs.LG].

[58] D. P. Kingma and M. Welling, *Auto-encoding variational bayes*, 2014. arXiv: 1312.6114 [stat.ML].

[59] I. Shafkat, *Intuitively understanding variational autoencoders*, Towards data science, blog, Feb. 2018. [Online]. Available: https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf.

[60] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(87)90125-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257.

[61] T. Thalhamer, "Data evaluation of the subsurface smoldering event at the bridgeton landfill," Hammer Consulting Service., Cameron Park, CA 95682, Tech. Rep., 2105. [Online]. Available: https://dnr.mo.gov/env/swmp/bridgeton/DataEvaFinal.pdf (visited on 06/04/2020).

[62] Missouri Department of Natural Resources, archived reports, "Bridgeton sanitary landfill," Missouri Department of Natural Resources, Tech. Rep., 2017. [Online]. Available: https://dnr.mo.gov/bridgeton/BridgetonSanitaryLandfillReports.htm (visited on 06/04/2020).

[63] H. Murakami and B. V. K. V. Kumar, "Efficient calculation of primary images from a set of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, no. 5, pp. 511–515, 1982. DOI: 10.1109/TPAMI.1982.4767295.

[64] L. Gao, J. Jiang, J. Liang, S. Wang, S. Yang, and Y. Qin, "Pca-based approach for video scene change detection on compressed video," *Electronics Letters*, vol. 42, pp. 1389–1390, 2006.

Video Links for South New Jersey Landfills Results

The following are the video links for the thermal state images, K-means clustering and Heat index for each landfill in a separate table.

Table A1

Deerfield landfill-Cumberland county results

Video file name	Link in YouTube
Deerfield TWP Cumberland index	https://youtu.be/4lj22hqtrYM
Deerfield TWP Cumberland kmeans	https://youtu.be/n_1OjCvHSp4
Deerfield TWP Cumberland_LST	https://youtu.be/LaNHwVRVy34

Table A2

South Harrison Township landfill-Cumberland county

Video file name	Link in YouTube
South Harrison TWP Gloucester index	https://youtu.be/15AYH_LY2dc
South Harrison TWP Gloucester kmeans	https://youtu.be/IUH-SaqHfck
South Harrison TWP Glouchester_LST	https://youtu.be/ttWF4nzAaUI

Table A3

Carney's Point Township landfill-Salem

Video file name	Link in YouTube
Carney's Point Township index	https://youtu.be/x7U0VL7wEDg
Carney's Point Township kmeans	https://youtu.be/z5Vc44_MkG0
Carney's Point Township_LST	https://youtu.be/U9BuO7SZfBo

Table A4

Commercial Township Landfill –Salem County

Video file name	Link in YouTube
Commercial Township index	https://youtu.be/tMWBV0hEzMc
Commercial Township kmeans	https://youtu.be/m1229kDjOLw
Commercial Township_LST	https://youtu.be/caKHOW9Lw4k

Table A5

Egg Harbor landfill – Atlantic county

Video file name	Link in YouTube
Egg Harbor Township Atlantic index	https://youtu.be/FordFCn6gjk
Egg Harbor Township Atlantic kmeans	https://youtu.be/QZ3KwF6KWIE
Egg Harbor Township Atlantic_LST	https://youtu.be/WZFBFBmyhkE

Table A6

Mannington Landfill-Salem County

Video file name	Link in YouTube
Mannington TWP Cumberland index	https://youtu.be/F0BUVnstvzA
Mannington TWP Cumberland kmeans	https://youtu.be/WSFH0BXI8xw
Mannington TWP Cumberland_LST	https://youtu.be/OSgMVNLsquQ

Table A7

Millville City Landfil–Cumberland County

Video file name	Link in YouTube
Millville City Cumberland index	https://youtu.be/wzyU68mtnoc
Millville City Cumberland kmeans	https://youtu.be/abKIw6u3j94
Millville City Cumberland_LST	https://youtu.be/Yla80EdyC3U

Table A8

Vineland City Landfill –Cumberland County

Video file name	Link in YouTube
Vineland City NW Cumberland index	https://youtu.be/KCHFFP_2dlU
Vineland City NW Cumberland kmeans	https://youtu.be/RuMzHHvvX5M
Vineland City NW Cumberland_LST	https://youtu.be/Hds32d-ArJc

Table A9

Woodbine Landfill - Cape May County

Video file name	Link in YouTube
Woodbine Borough Cape May index	https://youtu.be/RpV-OcFzpBs
Woodbine Borough Cape May Kmeans	https://youtu.be/QHdbYk6OC7M
Woodbine Borough Cape May_LST	https://youtu.be/mRN3T0_kjRk

85

Table A10

Other Landfills

Video file name	Link in YouTube
Vineland City NE Cumberland index	https://youtu.be/DJidIN1Qm8Q
Vineland City NE Cumberland kmeans	https://youtu.be/W2nuzQzF7ug
Vineland City NE Cumberland_LST	https://youtu.be/T9KWzQJ49Pc
Vineland City S Cumberland index	https://youtu.be/BkrxLu4aeSI
Vineland City S Cumberland kmeans	https://youtu.be/xj3vNb1J6Y4
Vineland City S Cumberland_LST	https://youtu.be/H1kdxer9g8g
Vineland City SE Cumberland index	https://youtu.be/3nYLCrqsGhE
Vineland City SE Cumberland kmeans	https://youtu.be/bx-zEPQkm4s
Vineland City SE Cumberland_LST	https://youtu.be/626X87mM0xI
Vineland City SW Cumberland index	https://youtu.be/ZbrP9OBgWCQ
Vineland City SW Cumberland kmeans	https://youtu.be/e8-46vfHWis
Vineland City SW Cumberland_LST	https://youtu.be/p6XjEEcOXm4

Dimensionality Reduction in Temporal Domain Used in chapter 4

B.1   Direct Cosine Transformation (DCT)

Each frame is converted to its Direct Cosine Transformation (DCT), from which a vector of DC coefficients corresponding to the original frame is formed as shown in Figure 1, where X_k is the vector form of the DC-coef. image.

$$DCT \longrightarrow \text{DC-Coeff} \longrightarrow X_k, \; X_k \in R^{(N \times 1)}.$$

where N is the dimension of the of X_k and k is the scene time index

B.2   Subspace Determination

The first M frames in the beginning of each video scene or after a change is detected, are used for subspace determination and consequently, to estimate the mean θ_0 and the covariance Σ, before the change.

$$\{X_k\}_{k=1}^{M} \Rightarrow P \in R^{(N \times M)}, M \ll N.$$

where P is the *data matrix* combing the fisrt (M) DC vectors of X_k in lexicographic order, N is the X_k vector dimension and M is the number of frames to estimate the mean before the change $\mu = \theta_0$.

B.3   More Dimensionality Reduction Using PCA

The PCA finds eigenvalues and eigenvectors of the *correlation matrix C* of size $(N \times N)$,

$$C = PP^T$$

87

Alternatively, we use the *implicit matrix* $(\tilde{C})$ of size $(M \times M)$, which is very much smaller in dimension compared to C [63, 64, 34],

$$\tilde{C} = P^T P$$

The PCA for $(\tilde{C})$ is computed and the result is the M largest eigenvalues $\tilde{\lambda}_i$ $(M \times 1)$, the diagonal elements of eigenvalues matrix and the eigenvectors $\tilde{e}_i$ $(M \times M)$. The M largest eigenvalues λ_i equivalent to the original correlation matrix C and the corresponding eigenvectors can be found using Equation 27:

$$
\begin{aligned}
\lambda_i &= \tilde{\lambda}_i \\
e_i &= \lambda_i^{(-\frac{1}{2})} P \tilde{e}_i
\end{aligned}
\tag{27}
$$

where $\tilde{\lambda}_i$ and $\tilde{e}_i$ in Equation 27 are the corresponding eigenvalues and eigenvectors of the implicit matrix $\tilde{C}$.

The eigenmatrix, $\Phi = \{e_i\}_{i=1}^M$ of size $(N \times M)$. Each new X_k is mapped to the subspace of eigenvectors corresponding to highest eigenvalues using Equation 28:

$$Y_k = \Phi^T X_k \tag{28}$$

Y_k is of size $(M \times 1)$, k is the scene time index index.

B.4   Estimation of the Mean and the Covariance Before the Change

The first $\{X_k\}_{k=1}^M$ vectors are mapped to the subspace of eigenvectors and yield the reduced dimensionality feature vectors $\{Y_1, \ldots, Y_M\}$, which will be used to estimate the mean $\theta_0 = \mu_0$ and the covariance Σ before the change. Thus, each new DC vector X_k will be reduced to a feature vector Y_k obtained by projecting X_k onto a subspace of eigenvectors corresponding to highest eigenvalues. That is, every new frame is represented by this feature vector. The change detection algorithm will sequentially input feature vectors until a

change is detected.

In this application we found that only one or two largest eigenvalues account for more than 96% of the total eigenvalues of the correlation matrix which will reduce the Y_k to a single scalar value or a vector of size (2×1) respectively.

B.5 Example for Estimating the Mean and Variance before the Change

In this work, each frame from the video file represents the original LST image, is reduced to a vector X_k of size (6864×1) as shown in Figure A1. Considering $M = 8$ in this example as the number of first frames needed to determine the eigenvectors subspace, the mean and variance before the change.

Figure A1

Dimensionality reduction using Discrete Cosine Transform followed by principle component analysis (PCA).

Total Reduction (Preprocessing) = (900 X 1200)/6864 ≈ 158 times

Formation of subspace:

The first $\{X_k\}_{k=1}^M$, $M = 8$ and $X_k \in R^{(6864 \times 1)}$, are used to form the data matrix P of size (6864×8).

$$P = [X_1, X_2, \ldots, X_8]$$

Then the correlation matrix C will be as follows:

$$C = PP^T,$$

$$(6864 \times 6864) = (6864 \times 8)(8 \times 6864)$$

and the implicit matrix $\tilde{C}$

89

$$\tilde{C} = P^T P$$

$$(8 \times 8) = (8 \times 6864)(6864 \times 8)$$

$$\Downarrow \text{PCA } (\tilde{C})$$

$$\lambda_i \text{ and } e_i \text{ using Equation 27}$$

$$\Downarrow \text{ Equation 28}$$

$$[Y_1, Y_2, \ldots, Y_8]$$

$$(8 \times 8)$$

$$\Downarrow \text{Estimate } \theta_0, \Sigma \text{ before the change}$$

$$\theta_0(1 \times 8),$$

$$\Sigma(8 \times 8)$$

Triggering the Detection Algorithm:

Each new X_{M+1}

$$\xrightarrow[\;(6864 \times 1)\;]{X_k} \text{ mapped to the subspace of eigenvectors } \xrightarrow[\;(8 \times 1)\;]{Y_k}$$

For the spatial detection, we divide each image into macroblocks of size $M \times M$, where M denotes the number of macroblocks in the spatial case and denotes the number of frames in the temporal case. Then the same steps C1-C4 were carried out for dimensionality reduction and estimating the mean and variance before the change.