Tourism Management 63 (2017) 179–186

Contents lists available at ScienceDirect

Tourism Management

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

Hydrocarbon pollution in underwater sinkholes of the Mexican Caribbean caused by tourism and asphalt: Historical data series and cluster analysis

José-A. León-Borges [a], M.A. Lizardi-Jiménez [b, *]

[a] Universidad de Quintana Roo, Unidad Cancún. Department of Computing Science, Cancún, Quintana Roo, Mexico
[b] CONACYT-Instituto Tecnológico Superior de Tierra Blanca, Av. Veracruz S/N Esq. Héroes de Puebla, Colonia Pemex. C.P, 95180, Tierra Blanca, Veracruz, Mexico

HIGHLIGHTS

- Hydrocarbon pollution in sinkholes of Quintana Roo, México from 2012 to 2015 is shown.
- Vacational seasons: "High" (July) and "Low" (October) affect hydrocarbon pollution.
- Anthracene, Benzo[*a*]pyrene and Pyrene were found and could be related to asphalt.
- Tourism contributes to hydrocarbon pollution in the sinkholes.
- Clustering can be used to observe and interpret the cause of pollution.

ARTICLE INFO

Article history:
Received 22 May 2017
Received in revised form
13 June 2017
Accepted 15 June 2017

Keywords:
Data series
Hydrocarbon pollution
Sinkholes
Tourism
Asphalt
Clustering

ABSTRACT

In this study historical data sets from 2012, 2013, 2014 and 2015 of hydrocarbon pollution in sinkholes of Quintana Roo are presented for two vacational seasons: "High" (July) and "Low" (October). Anthracene, Benzo[*a*]pyrene and Pyrene were found to be the main hydrocarbon contaminants and these poly aromatics hydrocarbons are strongly related to pollution by asphalt. Tourism contributes to hydrocarbon pollution in the sinkholes of Quintana Roo in two main ways: runoff of fuel from cars and, principally, from annual asphalting. Additionally, in this paper we conduct a cluster analysis to classify the pollution grade of the sinkholes. Clustering can be used to observe and interpret the cause of vacational seasonal pollution produced by tourism.

1. Introduction

Quintana Roo state, in the Mexican Caribbean, has a karst relief, with typical karst systems caused by dissolution, known regionally as "cenotes" (from the Mayan word "dzonot" meaning hollow with water), which are the main source of freshwater on the peninsula (Cervantes-Martínez, Adrián, Mezeta-Barrera, Mariela, & Gutiérrez-Aguirre, Martha Angélica, 2009). There is little information as to the physical, chemical and biological properties of the cenotes of Quintana Roo; in fact, many systems are unexplored, especially in the southern state. Existing studies for the area have focused on taxonomic analyses of the freshwater fauna, as well as some geomorphological and geochemical descriptions (Schmitter-Soto et al., 2002). Recent work (Metcalfe et al., 2011) found hydrocarbon pollution in sinkholes, particularly polycyclic aromatic hydrocarbons (Lizardi-Jiménez, Leal-Bautista, Ordaz, & Reyna-Velarde, 2015). Other work (Medina-Moreno, Jiménez-González, Gutiérrez-Rojas, & Lizardi-Jiménez, 2014) found the presence of hydrocarbons in underwater sinkholes of Quintana Roo related to tourism. The presented explanation to the existence of hydrocarbons is the runoff from cars driven on the roads nearby (Metcalfe et al., 2011). This was confirmed by recent studies that detected the existence of

* Corresponding author.
E-mail addresses: jleon@uqroo.edu.mx (J.-A. León-Borges), chamarripas@yahoo.com.mx (M.A. Lizardi-Jiménez).

http://dx.doi.org/10.1016/j.tourman.2017.06.018
0261-5177/

hydrocarbons in underwater sinkholes in Playa del Carmen (Gold-Bouchout, Metcalfe, & Drouillard, 2009). All of those studies were punctual at the time. Scarce information is available about historic pollution data in underwater sinkholes in the Mexican Caribbean in general and Quintana Roo in particular. If gasoline runoff from cars is the main source of sinkhole hydrocarbon pollution then BTEX (benzene, toluene, ethyl benzene, and xylene) must be found. BTEX is the most soluble of the major gasoline compounds and, therefore, are common indicators of gasoline contamination (Wilson, & Moore, 1998). Additionally, if hydrocarbon pollution is related to asphalt or bitumen then the presence of polycyclic aromatic hydrocarbons (PAHs): Anthracene, Benzo [*a*]pyrene, Pyrene and Benz [*a*]anthracene in leachate coming off of the asphalt should be considered (Brantley, & Townsend, 1999). Cluster analysis or clustering has benefits when analysing hydrocarbon pollution in sinkholes caused by tourism. Additionally, is necessary to observe and interpret the cause of pollution produced by tourism, in terms of vacational season.

In this study we utilise four years worth of data from underwater sinkhole samples. About eighty samples are currently analysed but this number is still increasing and soon will be beyond analysis by hand.

Cluster analysis is used in many areas, including bioinformatics (Raza, 2012), biology, geology and chemistry (Lavine & Mirjankar, 2006) in order to find classes (group samples) when information is scarce.

Cluster analysis and data classification is mainly made using data mining techniques that employ k-means clustering. This algorithm is used in many different areas of science, including cluster analysis and classification of heart sounds (Amit, Gavriely, & Intrator, 2009), analysis of cyclone tracks (Yu, Zheng, Wu, Lin, & Gong, 2016), regionalisation of the European domain into regions according to climate change (Carvalho, Melo-Goncalves, Teixeira, & Rocha, 2016) and open cluster membership for coordinate positions of each star image in the galaxy (Aziz, Selim, & Essam, 2016). Data mining analysis data are stored electronically (Witten, Frank, & Hall, 2011) in databases. The classification task may be categorised as supervised or unsupervised. The unsupervised method is used when there is not a target variable; alternatively, a supervised method is employed when there is a pre-specified target variable (Larose & Larose, 2015).

We employ a basic technique called "k-means". This technique groups the data samples, thus segmenting a population into a number of clusters. Consist to create a cluster and add each sample within. Also a centroid is created per cluster and is used to measure the distances between sinkhole hydrocarbon pollution samples. Euclidean distance, the distance measure between two samples in a euclidean space, is used to obtain the distance between the centroids. Comparing each media value measure of distance its closer was as we decided that two measures belonged to the same cluster.

We observe two conditions: firstly, the separation in two or three clusters can show by which touristic pole the sinkhole is less/more contaminated by; second, the association of pollution for each sinkhole.

The aim of this work was to evaluate the hydrocarbon pollution in underwater sinkholes in the Mexican Caribbean over four years (2012–2015) and their correlation with tourism activity and asphalt pollution. Cluster analysis was used to analyse information on hydrocarbon pollution to help to categorise the sinkholes by groups of pollution. These groups can be two or three clusters. Two clusters when we want to know the least contaminated and most contaminated sinkhole. Three clusters for a low, medium and high level of pollution.

2. Materials and methods

2.1. Sampling points

Table 1 displays the name of tourism pole and the coordinates of each sample of the studied sinkholes in Quintana Roo: a total of 33 samples were taken each vacational season.

Samples were taken from the superior water body from a depth of between 1 and 1.5 m and put in amber glass bottles. Vessels were filled without air bubbles. Samples were labelled and kept at low temperatures during transportation to the laboratory (4 °C ± 2 °C). The complete sampling procedure was executed according to the Mexican normative (NMX-AA-014-1980).

Fig. 1 displays the studied sinkholes in Quintana Roo: Cancún City (A, B, C), (21°11′ 25.5″ N, 86°50′ 14.9 ″W; 21° 9′ 41.3″ N, 86° 51′ 0.81″ W; 21° 8′ 19″ N, 86° 51′ 39.35″ W), Puerto Morelos (D, E), (20° 50′ 53.7″ N, 86° 52′ 33.98″ W), Riviera Maya (F, G), (20°39′ 43.59″, 87°4′ 7.25″ W; 20° 12′ 31.26″ N, 87° 28′ 21.81″ W); Cozumel (H) (20°26′35″N, 86°59′40″W); Chetumal (I, J), (18° 46′ 2.37″ N, 88° 18′24.92″ W; 18°30′ 50.98″N-88°25′ 27.29″ W); Holbox (K), (21°32′ 16.0 ″ N, 87°13′ 12.0 ″ W). A total of 33 samples were taken each vacational season.

2.2. Hydrocarbon concentration evaluation

The presence of hydrocarbons in water samples was detected by Gas Cromatography (GC) as described elsewhere (Lizardi-Jiménez et al., 2015).

2.3. Cluster analysis

We classified the historical data (2012, 2013, 2014 and 2015) of the hydrocarbons pollution in the sinkholes by using data mining. The classification can be done via clustering and is used to perform the "Cluster Analysis". We employed a basic technique of clustering that measures the "distances" between hydrocarbon pollution quantities of the sinkholes. This clustering technique is called "k-means" and it groups the data samples, thus segmenting a population into a number of clusters. The clusters are object groups that internally and automatically find classes. Each class is a sample of water in each underwater sinkhole.

First of all, we must to understand the problem and data. Next we prepared the data. We made a dataset that contain every sample subtract of the underwater sinkholes. Second, we employ the algorithm with Matlab and finally graphed and observe the results for a better understanding of pollution sinkholes areas.

Our methodology for data analysis or cluster analysis is based on a generic model according to the comparison of data mining processes (Kurgan, 2006). The data mining generic process is: Application Domain Understanding (1), Data Understanding (2), Data Preparation and identification of Data Mining (DM) Technology (3), DM (4), Evaluation (5), Knowledge Consolidation and Deployment (6). The process is similar to Cross-Industry Standard Process for Data Mining (CRISP-DM) used by industry members (Larose & Larose, 2015).

The application domain (1) involves the basic idea to explore the hydrocarbon pollution underwater sinkholes in different areas to determine if the sinkholes share similar rates of pollution. We observe the data from Tables [1], [2], [3] and [4] for Data Understanding (2) and create a header file or nominal variables to be used in Weka and Matlab, We selected cluster analysis which employs k-means and hierarchical clustering. We focused on the k-means algorithm for DM phase (3) and (4).