Statistical Analysis for Automated Seep Extraction in GIS

Pennsylvania State University

Pamela Kanu, GISP
Academic Advisor: Patrick Kennelly
GEOG 586B: SUMMER 2017
Final Project
9/7/2017

Table of Contents

List of Figures and Table

Abstract

Sonar data are collected by acoustic beams that reflect off of the seafloor. Water column data uses the same technique, only data are collected in the water column of reflections from marine life or natural oceanic processes. The acoustic reflection data are imported into a text file where millions of points are used to create a representation of the seafloor. Water column data are used to map physical and biological characteristics of the ocean. Methods exist for extracting seeps from these massive datasets, however challenges must be addressed for their automation, web development and use in Esri's platform. In this analysis, millions of points conglomerate into dense masses. Multiple criteria decision analysis using ArcGIS Spatial Analysis tools—Point Density, Reclassify, Data Management Tools—convert, create, populate and modify fields, and Spatial Statistics—Hot Spot Analysis can identify the plumes and extract them. The results expose a coalesce of point clustering in a linear fashion seeping from the seafloor in a GIS Mashup that is automated in Python, then brought into a custom web application built with ArcGIS API for JavaScript 4.3 allowing the data to be distributed and effortlessly viewed in 3D.

1. Introduction

Seeps are natural gas and crude oil that enter the ocean through cracks and sediments along the seafloor. Seeps emit continuously into the ocean and drift with currents, analogous to an oil spill. Due to the warm environment caused by subsurface geologic activity, the escaping methane gas produce bacteria benthos consume (Wall, Jech, & McLean, 2016).
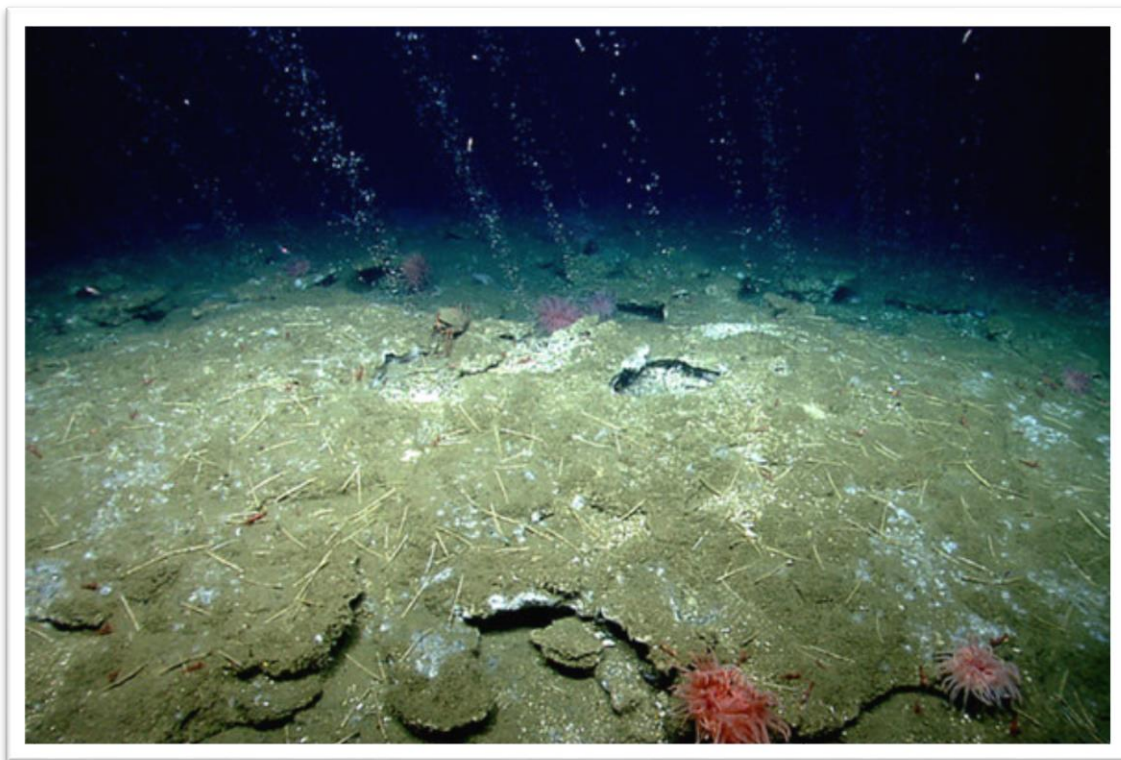
Figure 1. Natural seeps emitting from the seafloor (Woods Hole Oceanographic Institution, 2014)

The seep process is illustrated in Figure 2. Seeps form by oil escaping from the subsurface into the water column. Some seeps dissipate into the atmosphere, while the remainder

become an oil slick on the sea surface. Oil in the water column travels from the ocean surface back to the seafloor producing a fallout plume. National Oceanic and Atmospheric Administration (NOAA) is interested in oil spills to protect marine life and environment. Natural oil emissions can be toxic to birds and marine wildlife, damage surface waters and shorelines, which can impact human activity. NOAA's Restoration and Response team cleans areas of natural oil seeps and has estimated approximately 160,000 tons of petroleum enter North American waters naturally each year. The natural oil behaves much like an oil spill, both with viscosity ranging from motor oil to tar (NOAA, 2017).
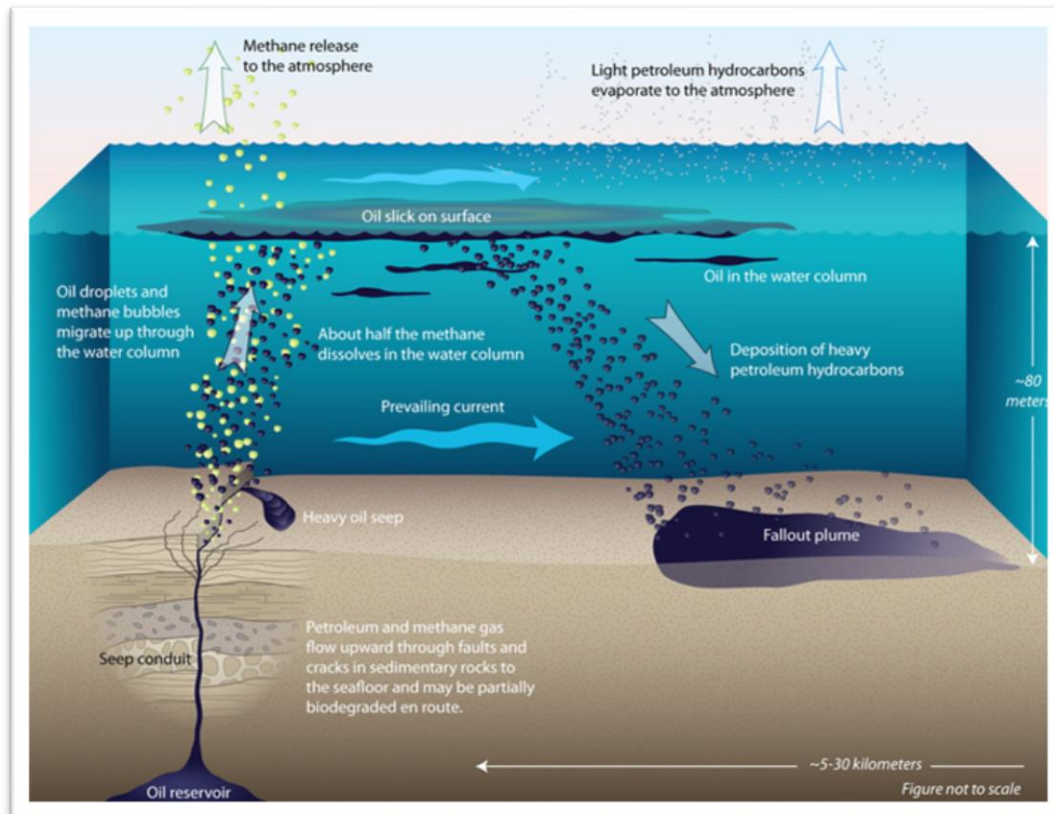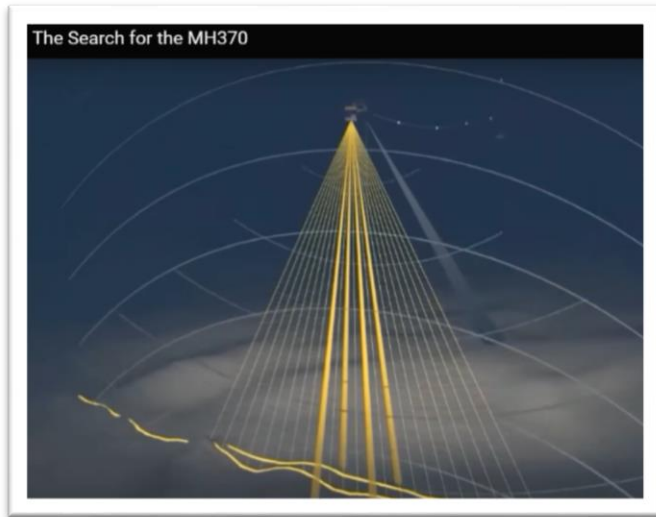

Figure 2. Natural Seep Process (Woods Hole Oceanographic Institution, 2014)

1.1 Data Acquisition

Data are acquired by sonar beams casted in straight-lines in a fan shape capturing meters of the surveyed area using GPS horizontal positioning and local water-level datums for vertical positioning. Precise horizontal and vertical positioning of the vessel are required to obtain an accurate sample of the water depth, water column, and seafloor. 4 dimensions (XYZ and Time) are used to calculate and correct sounding locations (Brown & Nell, 2003).

Figure 3. Data acquisition (Fugro, 2016)

1.2 Multi-Beam Echo Sounder (MBES)

MBES is sonar that emit sound waves beneath the ship's hull and uses spatial filtering (beamforming), a signal processing technique for directional signal transmission. Water depth is calculated by the time it takes the signal to reach the seafloor and reflect to the vessel.

For the scope of this project, bathymetric XYZ data are converted into a continuous raster and overlain slope to display the seafloor and identified seeps. The seafloor is estimated using multibeam echo sounder data at an average depth of approximately 84 meters.

Figure 4. Data acquisition using towing technique. Image courtesy of NOAA.

1.3 Water Column Data

Water column data are collected alongside MBES. The sonar captures an acoustic reflection of objects in the water column. The recorded data are used to map natural phenomena, marine life and seeps by calculating the time the sound to reflects to the sonar.

2. Industry

Existing seep extraction techniques were developed by University of New Hampshire Center for Coastal and Ocean Mapping Joint Hydrographic Center (UNH CCOM-JHC). The limitations for the software are the programs require a level of expertise to use them. These

programs lack the ability to automate tasks and routines with opened-source languages such as Python.

2.1 Fledermaus FMMidwater Tool

Fledermaus v7.7 FMMidwater tool is a software designed to rapidly view and extract water column features from a range of sonar file formats. Raw water column format are converted into a Generic Water Column format (GWC) to use for further processing and visualization by threshold filtering. FMMidwater provides the ability to export visualization objects and text files. The four steps for manipulating water column data include:

- Import sonar file(s)
- Convert to GWC format
- Identify features of interest using thresholds including:
  Depth Threshold Difference, where suspect soundings are selected based on a depth entered into a text field to filter all soundings more than the desired depth threshold away from the average depth in a bin
  Area-based filter, where the standard deviated threshold for each cell is specified
- Export to Fledermaus for visualization or ASCII for custom processing (QPS B.V., 2016)

The latest tools for improved feature detection utilized this architecture to transfer research code developed into the FMMidwater application. These tools incorporate techniques to help identify and visualize gaseous seafloor seeps (Heffron, McKenna, Doucet, Paton, & Beaudoin, 2014).

2.2 Qimera MidWater Tool

Qimera includes add-on packages for water column processing. UNH CCOM-JHC ceased Fledermaus development and extended the FMMidWater Tool in Qimera by further reducing sonar noise and converting the water column data sample into bathymetric format (QPS, n.d.).

2.3 Methane Flux

Once an area detects active seeps, the amount of methane flux expelled in an area can be estimated using quantitative methods. The methane flux uses multi-beam echo sounder and split beam echo sounder to measure target strength and estimate gas bubble size and velocity rise by means of a remotely operated vehicle (ROV). While this project does not encompass this type of detailed analysis, measuring methane flux would be suitable after seeps have been detected (Weber et al, 2014).

2.4 Using GIS for Data Management and Analysis

Employing multiple software for data processing is common in geological mapping. Geographic Information Systems provided a centralized platform for the creation of georeferenced and ancillary datasets to be viewed, analysed, mapped and managed (Orange, 1999). Esri ArcGIS10.5 product suite was used to integrate platforms—Python 2.7, Fledermaus, ArcGIS and JavaScript to provide data management for the processing, analysis and viewing of water column data. Python scripts were written to automate processing using data output from Fledermaus. ArcPy modules were used within Python to automate analysis

and apply spatial statistics. ArcGIS for JavaScript 4.2 provided the application interface for seep viewing.

2.5 Challenges in Seep Detection

2.5.1 Major Issues

The major hindrance with water column data is the substantial amount of information captured, causing data processing to be challenging. Automated methods for processing have yet to be developed. Advancement in research at UNH CCOM-JHC have led to a quick way to visualize the data, but no direct methods for seep detection using water column data in ArcGIS have been established (Gee et al, 2012). Currently, mapping seeps in ArcGIS are performed in the 2D, where manual extractions include digitizing a polygon around an area of seeps. Also, in 2D the data are limited to the XY location and not the Z. While ArcGIS has 3D capabilities, ArcScene pose challenges because the software is in a 32-bit environment and the newer 64-bit program ArcPro manifests constraints on the amount of computer memory and graphics used by the program.

2.5.2 Future Issues

Fledermaus and Qimera are optimal for processing raw data and appropriate for viewing seeps. These programs have limitations, as all parties needs to have the suitable software and proficiency to manipulate the data. Challenges remain with little automation for detecting seeps and maximizing computer memory is problematic, causing prevention in web development.

3. Capstone Objective and Study Area

My capstone objective was to map seafloor seeps for potential exploration site investigation and define a level of accuracy in seep extraction.

Bathymetric and water column data used in this analysis are favorable for site investigation due to the geologic activity on and below the seafloor. Raw water column data were processed by Fledermaus FMMidwater. Because of processing limitations, a subset area was extracted attributable to the data quantity. The study area is located in the Northern Atlantic Ocean.

4. Methodology

The process for automating seeps began with scripting data conversions in Python. The seep information is stored in raw TXT files and were converted to points using ArcPy modules. Rasters were calculated to display the seafloor and compute the density of the points in the water column. Both rasters were used to determine the weighted class by calculating water column heights and associating characteristics. After the characteristics are weighted, the Getis-Ord GI* equation is applied to further refine the results.

Step 1 Convert TXT files to points using ArcGIS	Step 2 Calculate Bathymetry and Compactness of Water Column Points	Step 3 Calculate Delta Z	Step 4 Define Criteria using Characteristics of a Seep	Step 5 Determine Weighted Class	Step 6 Getis-Ord GI* Hotspot Analysis

Figure 5. Methodology Overview

Step 1 – Data Conversions

Convert raw data to XYZ and XYZ to Points using Fledermaus FMMidwater and customized Python script.

The original water column data are provided in raw format as a .wcd (water column data) extension. The records were processed using Fledermaus FMMidwater tool. A point feature class was generated of the water column data. The bathymetry data converted into a raster using the *Point to Raster Tool.*

Step 2 – Calculate Derivative Rasters

Derivative rasters were created from the XYZ text files. Slope was calculated from the bathymetry to display the seafloor. The XYZ values were used to estimate the density using the *Point Density Tool* to measure the compactness of points in water column. From the observed data, the density tool estimates the concentrated areas by measuring the Z-value for the wcd points and compares them to the spatial relationship across the study area. The results produce continuous raster displaying density surface that is classified and converted into a polygon and incorporated into the analysis by location selections (Silverman, 1986).

Step 3 – Calculate Values for Seep Analysis

The bathymetry values were extracted and appended to the .wcd points and used to calculate water column heights, or Delta Z. The calculated values are used to remove noise by eliminating wcd point depths that fall below the seafloor.

Step 4 - Define Criteria for Seep Identification

Seeps were identified by multiple criteria classified based on the following characteristics:

4.1.1 Closer to the Seafloor (but not always the case)

Closer to the seafloor (but not always the case)

Seeps escape from the seafloor into the water column. Seeps form along cracks on the seafloor and thus tend to be linear features that are wider at their base. Water column heights are calculated by subtracting the MBES depth from the water column depth. Suggested by experts in the seep industry, a threshold of less than 4 meters in the water column were used to detect possible gas released closer to the seafloor.

4.1.2 Linear

Seeps appear linear

Components of a seep are naturally clustered. Wcd points are selected using Near_3D tool to find clustered points.

4.1.3 Condensed

Seeps are in dense
Seeps are dense in XYZ space. The XYZ points are used to calculate density to find the compactness of the points along the seafloor, but are also condensed in the through the water column depth. Density rasters were converted to polygons and intersected with water column points to isolate dense areas.

4.1.4 Disparate Amplitude Intensity

Amplitude values vary
Items in the water column reflect varying intensities, but related items reflect alike intensities. Amplitude values are relative to the survey. This analysis should not be combined with two separate projects; the results will be skewed.

Step 5 – Weight Classes

Thresholds and weighted classes were developed to classify the water column points. Each criterion came from characteristics described by in section 4 by experts in the industry. Each criterion was assigned a number from 5 classes—proximity to the seafloor, linear, density, amplitude intensity and noise was used to help detect seeps. The weighted classes were filtered to remove noise and classifications that register as insignificant to the recognized characteristics.

Step 6 – Getis-Ord Gi* Equation to Calculate Hot Spots

The Hot Spot Analysis tool was used to detect concentrations of high or low values (O'Sullivan & Unwin, 2010). Figure 8 displays the statistical equation used by the Esri *ArcGIS Hotspot Analysis Tool.* Given the weighted features, this tool identified statistically significant hot spots and cold spots using the Getis-Ord Gi* statistic. In this project, positive z-scores detected seeps, indicating the more intense the clustering of high values. The p-value determined the significance of the results. The Getis-Ord Gi* equation measures the association of weighted points determined in the methodology. A distance of 3 meters in the Z-value was applied to calculate statistics of neighboring values. This allowed for a high concentration of point densities to be confined in clustered areas   (Getis, 1992).

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum\limits_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum\limits_{j=1}^{n} w_{i,j}}{S \sqrt{\dfrac{\left[n \sum\limits_{j=1}^{n} w_{i,j}^2 - \left(\sum\limits_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \tag{1}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum\limits_{j=1}^{n} x_j}{n} \tag{2}$$

$$S = \sqrt{\frac{\sum\limits_{j=1}^{n} x_j^2}{n} - \left(\bar{X} \right)^2} \tag{3}$$

The G_i^* statistic is a z-score so no further calculations are required.

Figure 6. How Hot Spot Analysis: Getis-Ord Gi (Spatial Statistics) Works*

5. Results and Deliverables

5.1 Seep Extraction

The results produced a statistically significant feature class of extracted seeps with accompanying attributes retrieved during data acquisition. The water column data are viewed as 3D spheres in ArcScene with the bathymetry layer overlain the slope layer. For viewing the seeps via the web, the data are viewed in a custom web application built in ArcGIS for JavaScript 4.2. Seep information are viewed and distributed on more than one computer without commercial products such as ArcScene or Fledermaus.

The study area resides in the north Atlantic Ocean with an average water depth of approximately 84 m. The example dataset display wcd as colored spheres of about 1 – 1.5 cm spacing. Figure 7 and 8 demonstrates the identification of linear seepage expelled from the seafloor into the water column. Figure 9 displays noise and misclassified seeps. It is important to note the quality of data is critical and greatly impacts results of automated detection methods. The data must be examined and corrected manually. This development is ongoing and continually updated to improve seep detection and remove noise (Wall, Jech, & McLean, 2016). The seep detection classified non-seeps at a 98% accuracy and seeps at an 81% accuracy. The results are further explained in Section 5.2 Accuracy Assessment.
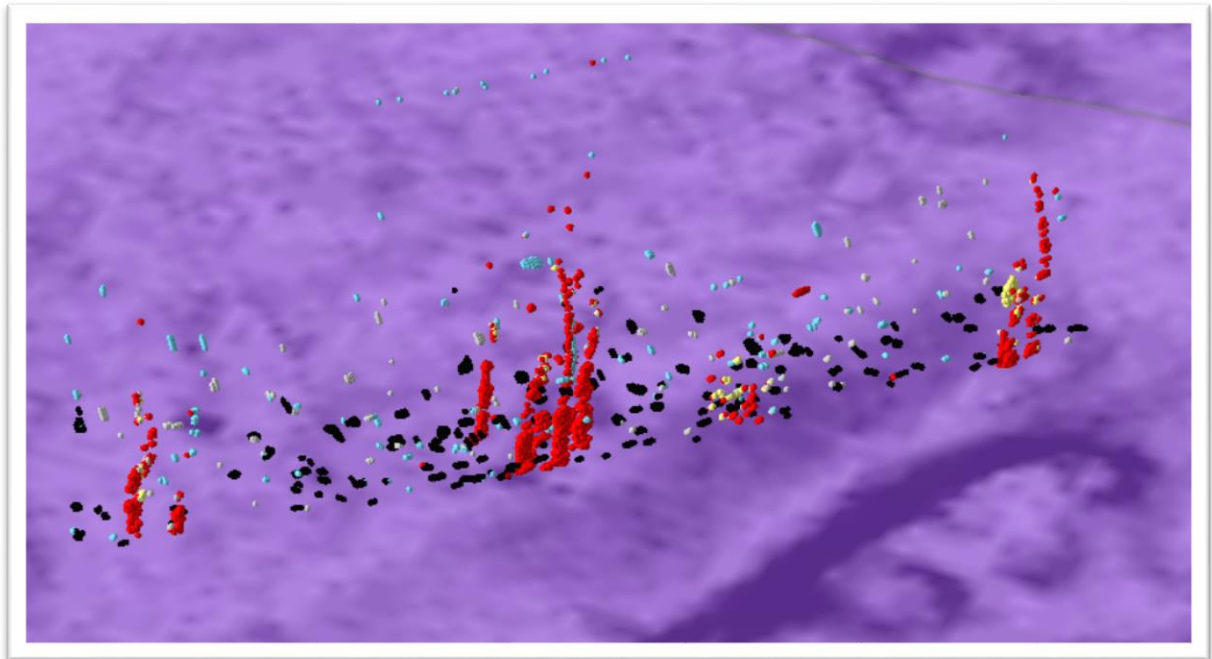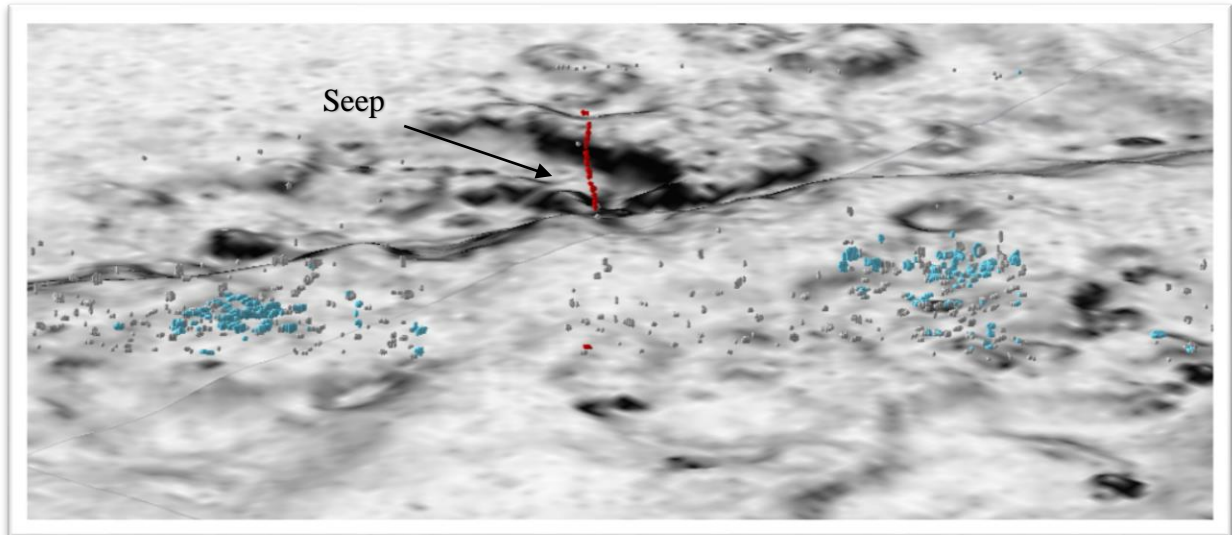

Figure 7. Seep Detection
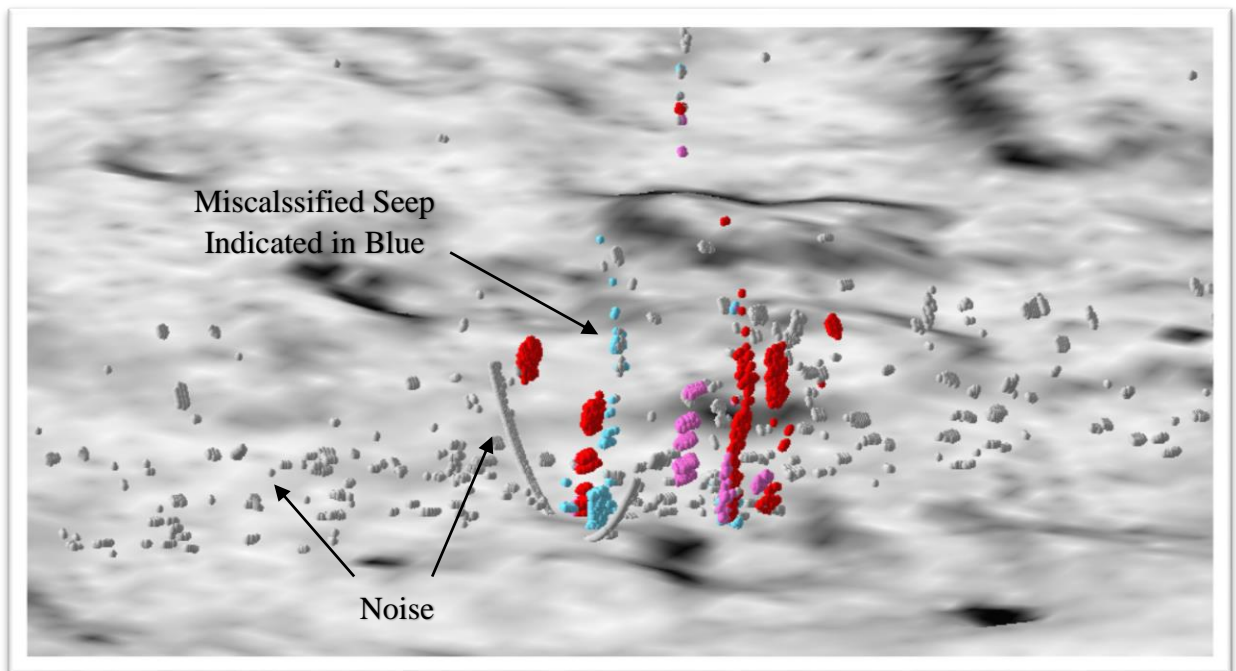
Figure 8. Linear Seep

Figure 9. Noise and Misclassification Areas

The automated processes for seep extraction were programmed in Python 2.7. Ascribable to the quantity of the dataset, the data were processes in Python's Integrated Development Environment (IDLE) with an AMD 64-bit processor on Windows 10.

5.2 Accuracy Assessment

100 random points were selected using the *Create Random Points* tool in ArcGIS of the statistically significant wcd points. Each of the points were ground truthed using an estimated sample of the seafloor, water depth and water column depths in a 3D environment. Each of the randomly selected points were identified as a seep or non-seep of the interpreted sample dataset. The results are totaled in Table 1. Errors of commission are labeled as No. Incorrect, where the analysis misclassified seeps. It is important to note producer's accuracy, as interpreted data are subjective (Esri, n.d.).

CLASSIFICATION	Total Random	No. Correct	Percentage Correct	No. Incorrect	Percentage Incorrect
NOT A SEEP	58	57	98.28%	1	1.72%
SEEP	42	34	80.95%	8	19.05%
TOTALS	100	91	91.00%	9	9.00%

Table 1. Accuracy Assessment Matrix

5.3 Future Development and Conclusions

The future development for this seep hunting analysis includes data processing advancements by exploring facets of computer programming. Using an advanced program for scripting and a computer with more power can optimize data processing for this analysis.
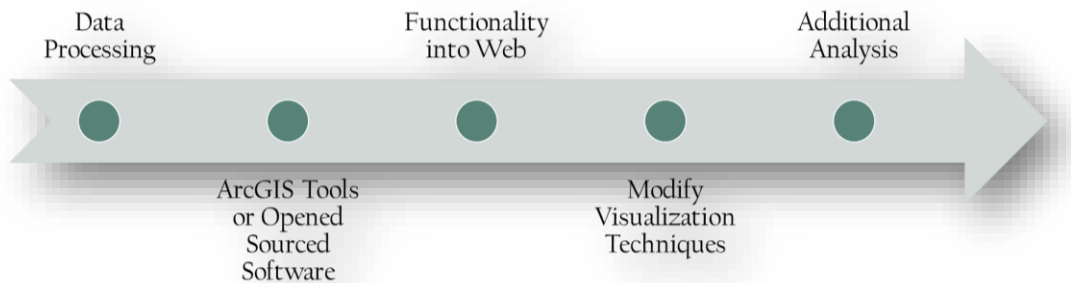

Figure 10. Future for Seep Detection Analysis

5.3.1 Data Processing and Development

To optimize the amount of time and memory used by Python and ArcGIS, further development for data processing to converting TXT files to points should be refined. Writing of analysis scripts and any intermediate processing required to implement analysis for the programming of the methodology, weighted criteria and spatial statistics could be improved. Multithreading, where RAM and graphic processing units (GPU) are in use simultaneously to optimize the scripting and automation. Moreover, executing multiple tasks at once can be achieved with Anaconda, a more powerful version of Python.

5.3.2 Functionality into Web

The web application was developed using JavaScript with ArcGIS for JavaScript API 4.2. The application could be used to process data via the cloud. Additional functionality should be programmed and tested for feasibility. In addition, data migration should account for fees and time to transfer files for web processing.

5.3.3 Visualization Modification

Visualization can be modified to enhance the viewing of seeps on a standard computer's memory or graphic card. This could be achieved by converting the points to more sustainable methods for ArcGIS. Such methods include a raster, aggregated hexagon, or triangular network, as performed in the results section for quality control.

5.3.4 Additional Analysis

Advancement for this analysis involves calculating the volume of seepage expelled in an area as displayed using a methane flux analysis as described in section 2.3 in this report. Additionally, predicting where gas plumes may travel through the water column at a predefined time or location and adding machine learning algorithms to teach the computer how to detect seeps may provide significant improvements to this analysis and the emerging seep industry.

References

Brown, J., & Noll, G. (2003). *MULTIBEAM SONAR DATA ACQUISITION SYSTEMS: A SIMIFIED CONCEPTUAL MODEL.* NOAA, U.S. DEPARTMENT OF COMMERCE Office of Coast Survey. Retrieved from https://www.nauticalcharts.noaa.gov/csdl/publications/TM_NOS-CS03_FY2003-Brown_SonarDataSystem.pdf

Esri. (n.d.). Retrieved from Accuracy Assessment for Image Classification: http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/accuracy-assessment-for-image-classification.htm

Fugro. (2016, October 01). *Search for MH370*. Retrieved from Fugro.com: https://www.fugro.com/media-centre/fugro-world/article/facts-of-MH370-search-vessel

Gee, L., Doucet, M., Parkerm, D., Weber, T., & Beaudoin, J. (n.d.). *Is Multibeam Water Column Data Really Worth the Disk Space?* Retrieved from http://www.qps.nl/download/attachments/14778902/Is+multibeam+water+column+data+really+worth+the+disk+space.pdf?version=1&modificationDate=1420018938603

Getis, J. O. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis, 24*(3), 189-206. doi:10.1111/j.1538-4632.1992.tb00261.x

Heffron, E., McKenna, L., Doucet, M., Paton, M., & Beaudoin, J. (2014, October). New Tools for Water Column Feature Detection, Extraction, and Analysis. Retrieved from http://www.qps.nl/download/attachments/14778902/New+Tools+for+Water+Column+Feature+Detection%2C+Extraction%2C+and+Analysis.pdf?version=1&modificationDate=1435656601120

Ingle, S. (2016, April 19). *WORLD'S LARGEST OFFSHORE SEEP-HUNTING SURVEY*. Retrieved from Fugro: https://www.fugro.com/media-centre/fugro-world/article/worlds-largest-offshore-seep-hunting-survey

Kongsberg. (2017, February). *30 KHZ MULTIBEAM ECHO SOUNDER*. Retrieved from km.kongsberg.com: https://www.km.kongsberg.com/ks/web/nokbg0397.nsf/AllWeb/A915A71E90B6CFAEC12571B1003FE84D/$file/306106_em_302_product_specification.pdf?OpenElement

NOAA. (2017, April 26). *Natural Seeps and Oil Spills*. Retrieved from Office of Response and Restoration: http://response.restoration.noaa.gov/oil-and-chemical-spills/oil-spills/resources/natural-seeps-and-oil-spills.html

O'Sullivan, D., & Unwin, D. (2010). *Geographic Information Analysis.* Hoboken, New Jersey: 2nd John Wiley and Sons.

Orange, D. L. (1999). Using Seafloor Mapping (Bathymetry and Backscatter) and High Resolution Sub-Bottom Profiling for both Exploration and Production: Detecting Seeps, Mapping Geohazards, and Managing Data Overload with GIS. *Offshore Technology Conference*.

QPS B.V. (2016, October 18). Fledermaus 7.7.x Manual. SAAB.

QPS. (n.d.). *Qimera Midwater add-on*. Retrieved from http://www.qps.nl/display/qimera/Qimera_Midwater

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall.

Wall, C. C., Jech, J. M., & McLean, S. J. (2016, March 06). Increasing the accessibility of acoustic data through global access and imagery. *ICES J Mar Sci, 73*(8). doi:2093-2103

Weber, T. C., L. Mayer, K. J., Beaudoin, J., Rzhanov, Y., & Lovalvo, D. (2014). Acoustic estimates of methane gas flux from the seabed in a 6000 km2 region in the Northern Gulf of Mexico. *Geochem. Geophys. Geosyst*, 1911-1925. doi:10.1002/2014GC005271

Woods Hole Oceanographic Institution. (2014, July 28). *Natural Oil Seep.* Retrieved from http://www.whoi.edu/oil/natural-oil-seeps