International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v10i5.5554
***Article Received**: 28 March 2022 **Revised**: 10 April 2022 **Accepted**: 25 April 2022 **Publication**: 31 May 2022*

Python Implementation of intelligent system for Quality Control of Argo Floats using Alpha Convex Hull

B. Naga Varun
Department of Computer Science & Engineering,
VR Siddhartha Engineering College, Vijayawada,India,
e-mail: nagavarunbatina007@gmail.com

S. Vasavi
Department of Computer Science & Engineering,
VR Siddhartha Engineering College, Vijayawada, India,
e-mail: vasavi.movva@gmail.com

Sreya Basu
Department of Computer Science, Pondicherry University, India
e-mail: 19370051@pondiuni.ac.in

Abstract— In each research field, quality control is the crucial part in order to ensure that the data collected is trustworthy. Quality control management helps in detecting abnormal behavior, monitor the health of detecting instruments. Argo Float is an ocean instrument which gives a lot of information regarding the conditions of the ocean. The information obtained has its value only if the quality of the parameter measuring procedure involves good quality tools. So, maintaining the quality of the procedure is as important as analyzing the obtained data. The salinity and the temperature data from Argo floats which collect data from in situ regions of ocean need to be scrutinized for anomalous values and examined to find out the reason behind the abnormal values recorded using a new method proposed by using Alpha convex hull. The data collected is used to construct an alpha shape polygon which is fine tuned to obtain best possible results. This method segregates by representing good data inside the polygon and the bad data outside the polygon. The proposed intelligent system is observed to detect the anomalous data occurring due to various problems such as sensor corrosion, bias ,spikes etc. This method aims to perform quality control for oceanographic data of argos at present but aims to work on different sources. The main advantages of the proposed system can be seen as the ability to identify anomalous data from in situ bulk data in minimal steps and to be applicable for cumulative data of various depths together.

Keywords- Argo floats, Quality control, Oceanographic data, Alpha Convex Hull, anomalous data

I. INTRODUCTION

Oceans have attained their utmost significance in the recent years due to the rich resource of valuable information. Still, only a part of ocean is only explored and information is extracted. This data comes into use only if data is extracted with utmost care and quality is maintained thoroughly. The quality of sensors which collect information is of paramount importance to maintain the quality. Argo Float is an instrument which collects the salinity and temperature values remotely and communicates with a satellite to transfer the information. This instrument is advantageous over other methods due to its remote working unlike manual operating to perform the required task.

Argos started with an initial number of just 3000 across the oceans present in the world which could gather and transfer vital information collected across various depths ranging from 0 to 2000 m. The data from Argos need to be monitored at regular intervals to analyze the oceanic conditions and act accordingly. Argos generally possess a lifetime of 4 years but several issues or errors may occur in its functioning which are to be identified through this method and get rectified to avoid errors. The relation between temperature and salinity is utilized to here to construct an alpha convex hull in the region of interest and identify the outliers which need to be examined by experts to identify the root cause.

II. BACKGROUND

Quality control of Argo floats has been a research topic for ages and different methods give different results of varying efficiencies. Observational in situ data is taken and the anomalous data is classified as an outlier which is given

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v10i5.5554
***Article Received**: 28 March 2022 **Revised**: 10 April 2022 **Accepted**: 25 April 2022 **Publication**: 31 May 2022*

through various methods. These referred papers have been useful in giving an idea about author's perspective.

In the paper [1], the author proposed a method which includes considering various parameters such as pressure, depth, salinity and temperature values and plotting with two variables to obtain a relation. The data is a cumulative dataset of certain standard depths which eliminates the need to interpolate data for various depths separately thereby reducing the chance for error. World Ocean Atlas 2013 data along with observational data is utilized for ground truth and analyzing. The key approach followed is to identify the anomalous data as outliers by a new method using Alpha convex hull. Several types of issues covered here are biofouling, spikes, sensor issues etc. The authors are keen to work on refining this approach in order to be applicable for other oceanographic data.

The authors of paper [2] explore the use of the metrics to understand, analyze and evaluate oceanographic data. New measures as considered for developing a relation for research that include Quality data expected, Total Data Return that gives an estimate about total data given to Argos as input and Quality data returned. The return values of Quality data returned will be higher for temperature values due to biofouling problem than salinity values. These metrics are considered in order to analyze the ocean data and monitor the quality control of Argos using the in situ data from oceans.

The authors discussed in [3], the availability of oceanographic data for research purposes. The oceanographic data including the salinity and temperature values are predominantly taken from World Ocean Atlas. The climatology data include most of the data but the hydrographic data is not accurately present from Indian coastal region. The indigenous organizations extensively work on bringing good source of data and thereby providing with accurate data. This significantly improves the data in terms of enhancing the actual oceanographic data to be available. This has been possible by adding various seasonal, annual, monthly oceanographic data to make it accurate enough for indigenous research to take this data. This gives the advantage of much stable climatological data for research in Exclusive Economic Zone of India..

The authors in paper [4] described an approach in which they proved that the alpha convex hull is defined as a generalization of the convex hull. Here, by proposing an algorithm it is shown that the construction of an alpha convex hull is efficiently computable. Several methods can be followed to deduce the algorithms for convex hull construction. Ron Graham presented the first algorithm in 1972 to construct a convex hull for a set of points in a plane. The applications for convex hull are explained elaborately in this paper like Cover Designing, Fingerprint matching, Geographical Information Systems (GIS), Path Planning, etc. The alpha shape as a concrete geometrical concept of extending the concept of convex hull, in order to generate slightly different shapes based upon the measure of anti-clockwise angle alpha. The generalization of convex hull is performed in this paper to construct a simple n sided polygon for a planar set of points and introduced concept to convex hull and alpha shape in some applications.

III. METHODOLOGY

The proposed intelligent system as shown in Figure.1, identifies by using various algorithms and determines whether the float profiles lie within the convex hull and classifies them as good or bad profiles. In this research, convex hulls are built using World Ocean Atlas (WOA) 2013 climatology at a resolution of 0.25° x 0.25°. These alpha shapes constructed of various alpha values are then used to categorize good and bad in situ T/S data profiles. Quality tags are assigned to all types of concerns related with in situ oceanographic data. In this project all of these had been done in Python and all the important and necessary packages were used.

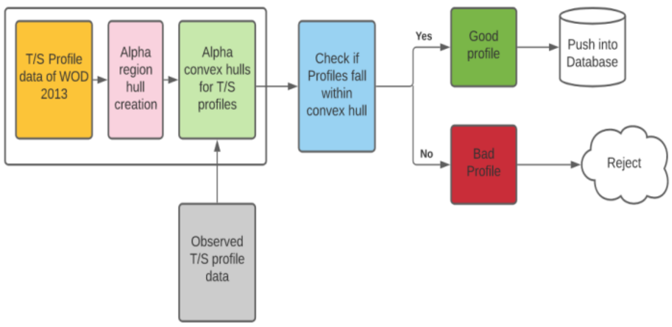

Figure 1. Architecture of the Intelligent System

For the implementation of this project, an approach is proposed which uses the concept of Alpha Convex Hull. Our approach deals with the processing of given input Argos data and outliers or bad data. The Point in Polygon principle is applied to the constructed alpha shapes using convex hull principle to differentiate among good and bad T/S profiles.

The approach contains the following modules:
1. Data Preprocessing - Data cleaning and Feature Engineering.
2. Alpha Region Hull construction
3. Alpha Convex hull construction for T/S profiles
4. Results are stored of alpha convex hull obtaining highest accuracy

A. Data Preprocessing - Data cleaning and Feature Engineering

• Data cleaning - In this module we will identify all the null or N/A values associated with the each attribute and will remove their corresponding rows as sufficient number of values for training and testing the model is available.
• Dropping the unnecessary features - In this process we will remove the attributes like `depth', `seasonal', `monthly', `ags' which are not useful for the project.

B. Alpha Region of Interest Construction

A region of interest (ROI) is created based on the path of observed Argo temperature and salinity profiles. ROI is also thought to be the most efficient convex hull (alpha shape), which includes the Argo float observations' route. The x-coordinate marks the longitude, and the y-coordinate is used to mark the latitude. The raw data of Argo floats was read from World Ocean Atlas 2013 with a 0.25° x 0.25° resolution. By building a convex hull that covers all places, the region covering all floats trajectories was retrieved.

C. Alpha Convex hull construction for T/S profiles

Temperature and salinity profile data from the World Ocean Atlas (2013) for this ROI are used, and a T/S convex hull with an acceptable alpha value is produced. All the temperature values in the region are taken along the y-axis and salinity values are taken along the x-axis. The T/S Alpha convex hull is built using standard deviation of temperature and salinity. If P(S,T) is the data with salinity and temperature, a union of lists with data containing :
P(S,T) , P(S+2*sd,T) , P(S-2*sd,T), P(S,T+2*sd) , P(S,T-2*sd).

All points lying inside the Alpha region hull are plotted in this list union. For these plotted points, an alpha convex hull with alpha=0.045 is constructed. Figure. 3. presents the Alpha convex hull with alpha0.045 constructed with salinity along x-axis and temperature along y-axis from the Argos data.

D. Applying the point in polygon principle to find outliers

Following that, the Ray Casting Algorithm is employed to determine whether the recorded temperature and salinity profile falls within or outside this climatological T/S convex hull..PIP is a spatial process that involves overlaying points from one feature dataset over the polygon of another to determine whether points are contained within the polygon. It returns a true value if the point lies within the polygon and a false value, otherwise.

```
For any point P(x,y)
If polygon contains P(x,y):
return true
Else
return false
```

E. Finding the most accurate alpha value by fine-tuning

The profiles data is plotted in such a way that all the points lying inside the polygon are given quality flag as good indicated by green color and other points are given quality flags as bad which are indicated by red for data falling outside the polygon's perimeter. The alpha value is fine tuned to attain at most accuracy using a confusion matrix by reducing the number of false positives and false negatives by identifying erroneous profile data. The required alpha value after fine tuning is set as standard value. Accuracy is given by:
Accuracy (All correct values / All values) = TruePositives + TrueNegatives / TruePositives + TrueNegatives + FalsePositives + FalseNegatives..

F. Data Collection

The suggested method makes use of float data numbered 2900782 from the year 2007 to 2012 of gridded climatology data from the World Ocean Atlas 2013, which has a resolution of 0.25° x 0.25°. The data is collected by Argos at 102 standard depth levels from 0 to 5500m, and at 0.25° salinity, temperature and other variables along with horizontal resolutions, annual, monthly, seasonal data of temperature, oxygen, and salinity are supplied. Temperature and salinity climatologies are available for specific locations of the worldwide ocean with adequate data coverage. The data is then used to create alpha forms, which are used to distinguish between good and bad data.

G. Software Requirements

- Alpha shape
- Shapely
- Numpy
- Pandas
- Scikit-learn
- Python 3.6

H. Hardware Requiremetns

- System – Intel Core i5
- Disk Space – 2 GB
- RAM – 4 GB

IV. RESULTS

A. Defining Alpha Region of Interest

To read the raw data of Argo floats, data from World Ocean Atlas 2013 which is at 0.25° x 0.25° resolution was used as shown in Figure 2(a), (b). The region covering all floats' trajectory was extracted by creating an α convex hull that covers all positions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 5
DOI: https://doi.org/10.17762/ijritcc.v10i5.5554
***Article Received**: 28 March 2022 **Revised**: 10 April 2022 **Accepted**: 25 April 2022 **Publication**: 31 May 2022*

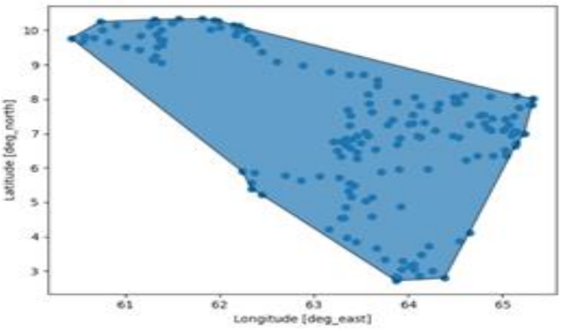

Figure.2 (a) Convex Hull

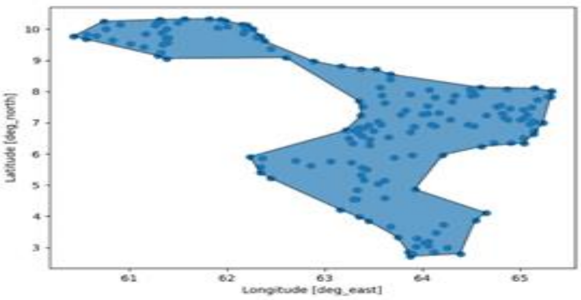

Figure. 2 (b) Alpha shape

B. Alpha Convex hull for T/S profiles

Figure.3 depicts a Salinity vs. Temperature convex hull constructed with a suitable alpha value utilizing temperature and salinity profile data from the Annual World Ocean Atlas (2013) for this Region of Interest.

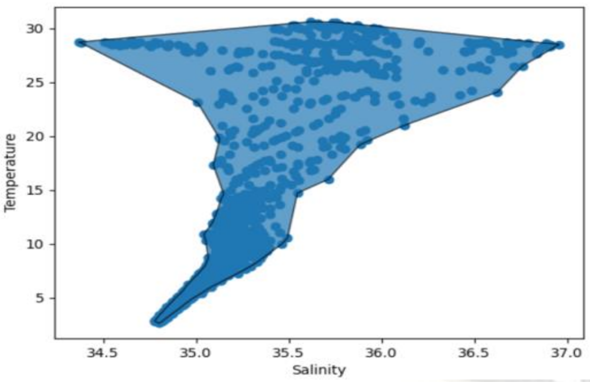

Figure.3. Alpha convex hull of Salinity vs. Temperature

C. Finding outliers

The results of the proposed system tested with different input values. Our proposed system reached an accuracy of 98.42% after fine tuning the alpha over various values and this occurred at alpha=0.045. In Figure. 4(a) that is the final output page which gives the outliers with maximum accuracy for alpha=0.045.

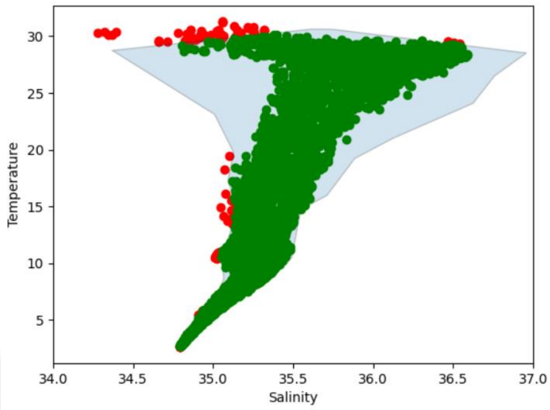

Figure. 4(a) The Profiles of Argos observed overlaid on T/S Convex hull for α=0.045)

The quality flags taken from the World Ocean Database are considered as good i.e '1' if the Annual Salinity and Annual Temperature flags are '1' and '0' for other three cases which is considered as the Actual values list. The quality flags obtained by implementing this methodology which give different values with various parameters and Point in Polygon principle are considered as observed values list. The confusion matrix is generated for various alpha values and fine tuning is performed to get the best accuracy and that particular alpha value is considered for all further deeds.

From the confusion matrix, four metrics measuring the correctness of our model can be calculated. Figure. 4 (b) gives the accuracy metrics of the final output, with the help of the confusion matrix. The accuracy is tested for various alpha values in order to find the best alpha value that minimizes false positives and true negatives.

i. Accuracy = TN+TP / (TN +TP+ FP + FN)
ii. Recall = TP / (FN+TP)
iii. Precision = TP / (FP + TP)
iv. F1-Measure = 2*Recall*Precision / Recall + Precision

For alpha=0.045, obtained accuracy = 98.42% , recall = 98.57 % , precision = 99.85% and F1-measure = 99.20 % .

We consider accuracy as the primary measure to obtain best output and hence alpha=0.045 is the required alpha value that gives the best output.

```
Confusion matrix :
 [[8077  117]
 [  12    5]]
Outcome values :
 8077 117 12 5
Accuracy :  0.9842893679210815
```

Figure. 4 (b) Confusion matrix(for α=0.045)

V. Conclusion and Future Work

Quality assurance or quality management are terms used to describe quality control. This method can also be used to find and detect the spike offsets in many forms of profile data collected. The capacity to handle large volumes of profile information from any oceanographic base is a major benefit of employing the convex hull approach. The profiles which are rejected using this method can be passed through expert visual evaluation making their task less complicated and much easier. .However, in order to obtain the finest possible convex hulls for performing outlier analysis, this approach necessitates that climatology be kept up to date at all times. Furthermore, this strategy can be used with other methodologies utilized by the oceanographic community to improve data study quality and application. Future study should focus on examining the suggested system's conclusions on a wider dataset from diverse ocean regions to ensure correctness. Future study should include a comparison of our model to other research and calculating the correctness of the proposed system.

References

[1] R Venkat Shesu, TVS Udaya Bhaskar, E Pattabhi Rama Rao, M Ravichandran, B Venkateswara Rao An improved method for quality control of in situ data from Argo floats using α convex hulls, MethodsX: Volume 8, 101337, 2021, https://doi.org/10.1016/j.mex.2021.101337

[2] Satish R U V N , T V S Udaya Bhaskar, Metrics for the assessment of quantity and quality of data by Argo floats, Indian journal of Geo-marine sciences 50(3):187-192

[3] Chatterjee, A., Shankar, D., Shenoi, S.S.C. et al. A new atlas of temperature and salinity for the North Indian Ocean. J Earth Syst Sci 121, 559–593 (2012). https://doi.org/10.1007/s12040-012-0191-9

[4] Saeed Asaeedi, Farzad Didehvar and Ali Mohades, Alpha Convex Hull, a Generalization of Convex Hull, Computational Geometry Article, arXiv:1309.7829, 2013.

[5] H. Edelsbrunner, D. Kirkpatrick and R. Seidel, On the shape of a set of points in the plane, IEEE Transactions on Information Theory, vol. 29, no. 4, pp. 551-559, July 1983, doi: 10.1109/TIT.1983.1056714.

[6] Climate Data Guide, World Ocean Atlas-2013[WOA] https://climatedataguide.ucar.edu/climate-data/world-ocean-atlas-2013-woa13 Last accessed on April 10th 2021.

[7] R.A Jarvis, On the identification of the convex hull of a finite set of points in the plane, Information Processing Letters 2, 1973,18-21, North Holland Publishing Company.

[8] Kai Petersen, Klaes Wohlin. The Waterfall model in a large scale development, International Conference on Product-Focused Software Process Improvement, 2009, DOI: 10.1007/978-3-642-02152-7_29,

[9] Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA; 2012

[10] Harris, C.R., Millman, K.J., van der Walt, S.J. et al, Array programming with NumPy,Nature 585, 357–362,2020.

[11] Herbert Edelsbrunner, Ernyst P. Muke, Three dimensional Alpha shapes, ACM Transactions on Graphics, October 1994,

[12] J. D. Hunter, Matplotlib: A 2D Graphics Environment, Computing in Science Engineering, vol. 9, no. 3, pp. 90-95, 2007.

[13] Joel Lawhead, Learning Geospatial Analysis with Python , 2013

[14] David Cournapeau ; Sci-kit learn: Google Summer of Code project, 2007

[15] Roger Pressman and Bruce Maxim - Software Engineering: A Practitioner's Approach, 9th Edition, McGraw Hill Higher Education.