Machine Learning Model to Predict Potential Fishing Zone

Swaroop Laxmi Mudliar*, Sai Shashank and M Chandak*****

***Abstract*- A key challenge today in aquatic environment conservation is the accurate tracking of the spatial distribution of various human impacts on activities like fishing. In the present paper an approach to identify the potential fishing zones in deep sea waters is developed using Autoregressive Integrated Moving Average(ARIMA) and Random Forest model. A large data set containing Indian fishing vessel track from 2017-2019 was taken as database. In the present paper an approach a methodology was developed to detect and map fishing activities. Validation of the model was done against expert label datasets which showed detection accuracy of 98 percent. Our study represents the first comprehensive approach to detect and Identify Potential fishing zones with the help of two important water quality indicators viz Dissolved Oxygen and Salinity.**

Key words - Water quality indicators, Satellite-based Automatic Information Systems (S-AIS), Datamining, Potential fishing zone (PFZ)

I. INTRODUCTION

Today coastal fisheries in national waters are closely monitored in some countries, however existing maps of fishing effort in Indian waters need to be more accurately track, especially in remote areas and the High Seas. Better understanding of the behavior of the global fishing fleets is required in order to prioritize and enforce fisheries management and conservation measures[1] worldwide. Satellite-based Automatic Information Systems (S-AIS) are now commonly installed on most ocean-going vessels and have been proposed as a novel tool to explore the movements of fishing fleets in near real time. A Data Mining (DM) approach using an algorithm inspired from studies on water quality indicators proves to give accurate prediction of PFZ. There are several water quality indicators out of which dissolved oxygen and salinity are important to identify PFZs[3]. For purse seiners a multi-layered filtering strategy based on vessel speed and operation time was implemented. Validation against expert-labeled datasets showed average detection accuracies of 83% for trawler and longline, and 97% for purse seiner. Our study represents the first comprehensive approach to detect and identify potential fishing behavior for three major gear types operating on a global scale. We hope that this work will enable new efforts to assess the spatial and temporal distribution of global fishing effort and make global fisheries activities transparent to ocean scientists, managers and the public. Dissolved oxygen (DO) is a crucial water quality parameter that influences the living conditions of all aquatic organisms that require oxygen. The level of DO in water bodies can be affected by anthropogenic activities and natural occurrences in catchments. Water temperature, the amount of oxygen taken out of the system by respiring and decaying organisms, and the amount of oxygen put back into the system by photosynthesizing plants, stream flow, and aeration are the factors that control the amount of dissolved oxygen in waterbodies. The water temperature highly influences the amount of DO; in other words, less oxygen dissolves in warm water than cold water.Different sensors mounted on satellites and other platforms, such as aeroplanes, measure the amount of radiation at various wavelengths reflected from the water's surface. These reflections can be used directly or indirectly to detect different water quality indicators, such as total suspended solids (TSS), chlorophyll-a concentration, turbidity, salinity, total phosphorus (TP), Secchi disk depth (SDD), Temperature, pH, Dissolved Organic Carbon (DOC),etc[10]. The spectral characteristics of water and pollutants, which are functions of the hydrological, biological and chemical characteristics of water, etc, are essential factors in the monitoring and assessment of water quality. The study thus introduces the widely employed spaceborne and airborne sensors in water quality investigations and discusses the utility of remotely sensed techniques in the qualitative assessment of waterbodies. Various properties (spectral, spatial and temporal, etc.) of spaceborne and airborne sensors are tabulated to be used as a sensor selection guide. Finally, based on the literature survey, the study presents a compilation of the various sensors useful in the study of some measurable water quality parameters, and investigates in more detail eleven water quality parameters based on the employed approaches to measuring their concentrations.Dissolved oxygen (DO) is a crucial water quality parameter that influences the living conditions of all aquatic organisms that require oxygen. The level of DO in water bodies can be affected by anthropogenic activities and natural occurrences in catchments. Water temperature, the amount of oxygen taken out of the system by respiring and decaying organisms, and the amount of oxygen put back into the system by photosynthesizing plants, stream flow, and aeration are the factors that control the amount of dissolved oxygen in water bodies. The water temperature highly influences the amount of DO; in other words, less oxygen dissolves in warm water than cold water.

*Department of Chemistry, **Department of Computer Sciience (shashanksai.ss@gmail for future correspondence), ***HoD Computer Science, Shri Ramdeobaba College of Engineering and Management, Nagpur-13

One pertinent catastrophe is that the oxygen may diminish to levels that are lethal for most fish and many aquatic insects. However, as the river re-aerates due to atmospheric mixing coupled with algal photosynthesis that adds oxygen to the water, the oxygen levels will slowly increase downstream.

Routine methods to measure COD are based on points, and have the time-consuming and laborious disadvantages in obtaining the distribution patterns so that it is difficult to reflect the status of whole region synchronously. This kind of point sampling methods may give accurate measurements, but they are time and money consuming. Further and most importantly, they cannot provide the real-time spatial overview that is necessary for the global assessment and monitoring of water quality .Satellite remote sensing may provide suitable ways to integrate aquatic data collected from traditional in situ measurements. Sea Surface Salinity (SSS) [5] Salinity and temperature are important factors to identify the density of seawater, and in turn, density is a critical component driving the currents in the oceans. Therefore, salinity is one of the key variables worth considering when monitoring and modeling the circulation in oceans. The role of ocean circulation in moderating the climate is crucial, and thus, sea surface salinity (SSS) is also critical to determine the global water balance, productivity forecast models, as well as evaporation rates. For example, when the salinity is relatively low, the mixed layer will be more stable, and the nutrient pump may be partially inhibited, possibly leading to reduced productivity or a delay in the onset of spring and autumn phytoplankton blooms [6]. Seasonal and inter-annual variability of sea surface salinity represent limitations on the hydrologic balance and coupled ocean-atmosphere climate models . Salinity plays a crucial role in the air-sea exchange of gases.A review of the available literature confirmed that salinity and dissolved oxygen are important parameters to identify to PFZ.

II. METHODOLOGY

Step1: Prediction of dissolved oxygen and salinity:

Using the previous year data we can predict the trend of dissolved oxygen and water. I have used Seasonal Autoregressive Integrated Moving Average(SARIMA).

SARIMA:

Autoregressive Integrated Moving Average(ARIMA), is one of the most used forecasting methods for univariate time series data forecasting. This method can handle data with trend but it is missing the component of season that is ARIMA model expects data neither be seasonal or has the seasonal component removed. Since, in fishing Season play an important role that's why I decided to use SARIMA instead of ARIMA.

There are two components of SARIMA which are as follows:

1) Trend Elements:

There are three trend elements that require configuration.

- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.

2) Seasonal Elements:

There are four seasonal elements that are not part of ARIMA that must be configured; they are:

- P: Seasonal autoregressive order.
- D: Seasonal difference order.
- Q: Seasonal moving average order.
- m: The number of time steps for a single seasonal period.

The notation for SARIMA model is specified as:

SARIMA(p,d,q)(P,D,Q)m

Where the specifically chosen hyperparameters for a model are specified; for example: **SARIMA(3,1,0)(1,1,0)12**

Code:

The data is taken from from copernicus GLOBAL OCEAN 1/12° PHYSICS ANALYSIS AND FORECAST UPDATED DAILY and GLOBAL OCEAN BIOGEOCHEMISTRY ANALYSIS AND FORECAST for the dates 7/11/17-7/11/2019 and the data looks like this for water Salinity and for dissolved oxygen:

```
In [23]: df[['time','so']].head()
Out[23]:
```

	time	so
0	2019-09-30 12:00:00	35.026093
1	2019-10-01 12:00:00	35.024567
2	2019-10-02 12:00:00	34.896390
3	2019-10-03 12:00:00	34.835354
4	2019-10-04 12:00:00	34.838406

```
In [5]: df[['time','o2']].head()
Out[5]:
```

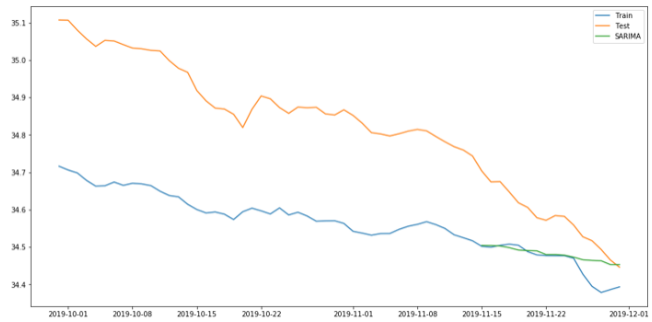
	time	o2
0	2017-12-05 12:00:00	207.75877
1	2017-12-06 12:00:00	207.76039
2	2017-12-07 12:00:00	207.86870
3	2017-12-08 12:00:00	207.81789
4	2017-12-09 12:00:00	207.31332

This the code for Water salinity prediction:

```
import statsmodels.api as sm
```

```
train, test = df[:int(len(df) * 0.75)], df[int(len(df) * 0.75):]
```

```
y_hat_avg = test.copy()
fit1 = sm.tsa.statespace.SARIMAX(train.so, order=(2, 1, 4),seasonal_order=(0,1,1
y_hat_avg['SARIMA'] = fit1.predict(start='2019-11-15', end='2019-11-30', dynamic
plt.figure(figsize=(16,8))
plt.plot( train['so'], label='Train')
plt.plot(test['so'], label='Test')
plt.plot(y_hat_avg['SARIMA'], label='SARIMA')
plt.legend(loc='best')
plt.show()
```



Step 2 : Prediction PFZ

I have used random forest algorithm for predicting Potential fishing zone. Random Forest consists of a large number of individual decision trees that operate as ensemble[10].Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is:

A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

```
In [234]: class_weights = {0.0:1, 1.0:50}
          clfRF = RandomForestClassifier(random_state=0,class_weight=class_weights)
          clfRF.fit(X_train, y_train)
          predRF = clfRF.predict(X_test)
          print(accuracy_score(predRF,y_test))

          /home/light/.local/lib/python3.6/site-packages/sklearn/ensemble/forest.py:248: FutureWarning:

          The default value of n_estimators will change from 10 in version 0.20 to 100 in 0.22.

          0.9835935955722475

In [236]: print(classification_report(y_test,predRF))

                        precision    recall  f1-score   support

                   0.0       1.00      0.98      0.99      7377
                   1.0       0.95      0.99      0.97      2741

             micro avg       0.98      0.98      0.98     10118
             macro avg       0.97      0.99      0.98     10118
          weighted avg       0.98      0.98      0.98     10118

In [227]: confusion_matrix(y_test,predRF)

Out[227]: array([[7227,  150],
                 [  16, 2725]])
```

III. Conclusion

We hope this work will enable new efforts to assess the water quality in terms of dissolved oxygen and salinity so that potential fishing zones are identified with the help of Autoregressive Integrated Moving Average and random forest models and help the coastal fishermen for better catch. This model is highly beneficial since it can be extended to several water quality indicators.

Supply of safe clean drinking water is vital to maintain public health. Effective means for ensuring water safety is through risk assessment and management approach starting from catchment to storage vessels at households. Using WSP approach, present study could identify risks factors affecting water quality throughout the distribution network. Intermittent water supply and poor maintenance of distribution network were identified as the major hazards causing entry of fecal coliforms in potable water. Minor risks were successfully addressed and immediate corrective measures were adopted. Be that as it may, this study suggests there is need to give a continuous assistance and guidance to local water management team for the maintainability of the WSP.

IV. References

1. Copernicus [Copernicus]

2. Incogis fishing data [incogis]

.