Scatterometer L1B Data Product Flag Verification using ML Technique

Dhairvi Shah Information and Communication Technology L.J Institute of Engineering & Technology

Scientist. Anuja Sharma SIPG/DQE Department SAC – ISRO Ahmedabad

Prof. Mitesh Thakkar H.O.D Information and Communication Technology L.J Institute of Engineering & Technology

*Abstract****: Scatsat-1 Scatterometer is a microwave non - imaging radar aimed to measure backscattered energy over the ocean surface. Backscattered energy helps in deriving wind vector – speed and direction at the spatial resolution of nearly 25 X 25 Km cell. Data received from satellite goes through a set of data processing steps and Level-1B data product is generated in HDF format. This data product includes Sigma-0 values, brightness temperature with precise location, Meta information, geometry related parameters and Sigma-0 flag for each footprint covered on Earth surface.***

***Data Quality Evaluation System is a system responsible to evaluate the quality of data product independently and generates quality metrics for the same. Apart from quality metrics, we can depict Sigma-0 flag that is an important parameter packed in data product from which node, beam, scan-direction, sigma-0 quality, sigma0 validity, sea-land boundary details, sea-ice information etc. are received. During data processing, this information is tagged with each footprint using various methods, criteria and references. Under data quality evaluation, we need to verify these flags using machine-learning techniques. It is planned to generate a model to identify the significant flag values using historic data from calibrated ScatSat-1 Scatterometer.***

*Index Terms****: Scatterometer, Level-1B, Backscattered energy, Sigma-0 flags, Data Quality Evaluation***

1. INTRODUCTION

Scatterometer is a non – imaging, pencil beam, active, microwave radar (**RA**dio **D**etection **A**nd **R**anging). This Ku band (13.65GHz) Scatterometer carries a parabolic reflector antenna, which measures back scatter strength using dual beams HH (inner beam) and VV (outer beam) covering 1400 km and 1800 km swaths respectively. With the above specifications, it covers full globe in 2 days and with these beams, they form 281 and 282 footprints respectively on earth. These footprints can further be divided into slices. This Data related to footprints and slices is present in Level-1B in HDF format i.e. Hierarchical Data Format used to store, manipulate and organize large amount of data, specifically scientific data.

The Level-1B contains physical parameters like Orbit Attitude Data (Roll, Pitch, Position/Velocity X, and Y and Z components, Yaw, Record Time) and Science Data (Kp, SNR, X-Factor, Range, Latitude, Longitude, Incidence angle, etc.) The Sigma-0 flag parameter provides flag values that helps in determining the quality of footprints or slices in terms of Scan direction (Aft/Fore), Polarization (HH/VV), Node (Descending/Ascending), Good / Poor sigma-0 / BT, Valid / Invalid Sigma-0 / BT etc.

Historic data from Scatsat-1 will train the model that will not only verify the flag values but also predict them for future phases of Scatsat-1 or upcoming missions like Oceansat-3

This paper helps in demonstrating the accuracy in predicting of flag bits using HDF files of the historic data for machine learning models. Section 2 provides the insight of working principle of Scatterometer as well as procedures for obtaining data product, which acts as an input for our model, whereas section 3 tells about the purpose of the model while approach is explained in section 4. Furthermore, each step of the approach is discussed in detail within section 4. Section 5 includes the results and analysis of proposed model marking the end of the paper.

1. Working Principle of Scatterometer

Scatsat-1 Scatterometer is an active radar that transmits chirp signal through HH (Inner) beam and VV (outer) beam covering 1400 Km and 1800 Km swaths respectively. It has a parabolic antenna that after sending the signal, measures the energy received from returning signal. This phenomenon of returning signal is known as backscattering giving us the backscattered energy. The following figure 1 shows insight of beams formed by Scatterometer.

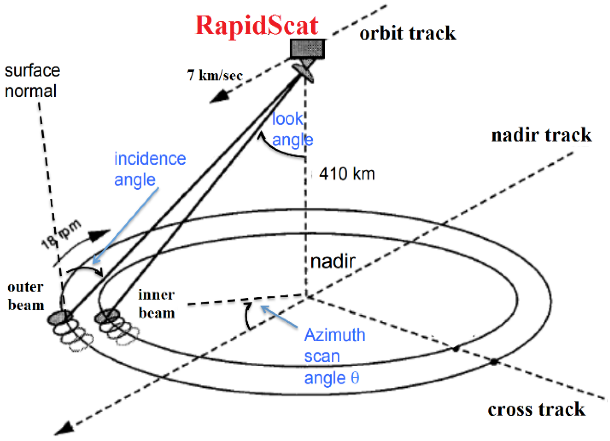


Fig 1 Inner/Outer Beams formed by Scatterometer

Effects of the signal from antenna is sent through antenna on periodic structure, if it is in range of ‘lambda’ (λ) i.e. it is equal to the value of *dsinθ* where *d* is the distance between two troughs, then it can be termed as Bragg scattering (Fig 2 ). Here, rougher the surface more will be the Bragg scattering resulting into higher backscattered energy. This increases the efficiency of the Scatterometer.

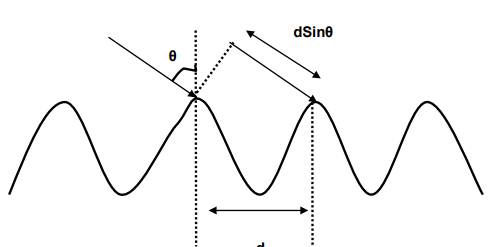


Fig 2 Bragg Scattering - Basic working Principle of Scatterometer

The backscattered energy received is the result of the strength of the Scatterometer. It contains the information regarding weather and the climatic conditions over the ocean surface. This data received is then sent to earth for its study and research purpose.

Resultant data is sent to the respective ground stations from Scatterometer. Data of all 14/15 orbits is being acquired daily in near real time from the ground stations. This data is geometrically and radio metrically corrected and packed and can be defined as data products which are synchronized for time and then it is separated **½** revolution wise. There are in total five levels of data products. They are Level 0, 1B, 2A, 3W and 3S . Level 0 is obtained from raw sensor data. Whereas each level after Level 0 is dependent on level preceding it. These levels of data products can be illustrated as shown in Fig3 given below.

Fig 3 Levels of Data Products

The proposed model uses the Level-1B data. Actual data is in Float form, which uses 4 bytes, because of which calculations like subtracting offset values and dividing that with scale values are conducted and float values are thus converted to short consuming 2 bytes only. They are available in HDF i.e. .h5 format. For our model, this level-1B .h5 file acts as an input data.

1. Purpose

Model aims to predict the sigma0 flag bits with the historic data obtained from Scatsat-1 Scatterometer. Application of Machine Learning algorithms on data formed from level-1B helps in fulfilling this purpose. The flag bits predicted by this model are:

1. South-Pole to North-Pole i.e. Ascending node (S-N)
2. North-Pole to South-Pole i.e. Descending node (N-S)
3. Aft scan-direction
4. Fore scan-direction
5. Good/Poor sigma0 or brightness temperature bit
6. Valid/Invalid sigma0 or brightness temperature bit
7. Land-Sea boundary

Sigma0 parameter that is obtained from Level-1B is the backscattered co-efficient that represents the strength of Scatterometer signal reflected by the distributed scatterer. This Sigma0 can be imperfect as its estimation can be noisy due to the instrument thermal noise and radar signal fading effects. Sigma0 flag bits tell about the quality of this Sigma0 for specific slices and footprints on earth. As a result, this bit information is very important because when this Data Product goes to the user, he should be informed to use it cautiously as there can be a low quality data for land, sea-ice and heavy rain conditions that has to be rejected.

From the above list, first five provides with information regarding satellite pass and viewing geometry of the antenna while remaining provides with quality information. Therefore, we can say that the main purpose of the proposed model is to provide us with the geometric and quality information by predicting the above flag bits.

1. Approach

The process for obtaining required model can be divided into 4 phases.

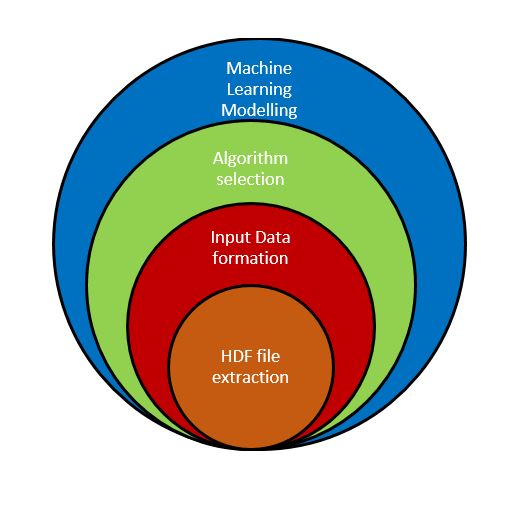


Fig 4 Approach of the Model

**Phase1: HDF File Extraction**

Extraction of this file provides us with needed physical parameters along with Scale and offset values or say information on attributes are obtained from this file.

**Phase2: Input Data Formation**

CSV file having bit wise data from the flag file along with some related physical parameter acts as the input data for further steps.

**Phase3: Algorithm Selection**

Machine learning is the process of meta-learning. Its main aim is to find the algorithm, which will have minimum error rate for the present dataset.

**Phase4: Machine-Learning Modelling**

For execution of any machine learning technique, there are two main requirements:

1. Data and
2. Algorithm that applies on that data

Need of the data is fulfilled by step 2 and that of algorithm by step 3. As a result, application of ML technique is possible. Model is trained with 80-90% of the data from the dataset while tested with remaining 10-20% of the dataset.

Each phase in detail can be explained as below:

Phase-I HDF File Extraction

As our proposed model is based on Level-1B data that is present in HDF format, the first and foremost step is to extract those data. As stated earlier 4byte, float data is converted to 2byte short to form Hdf file. As a result, while extracting data they should be calculated again using scale and offset values to get original values.

The original data is obtained back using same scale and offset values, which were used to convert it in the first place. Given below formula converts short values into float again.

*Scale\*data + offset*

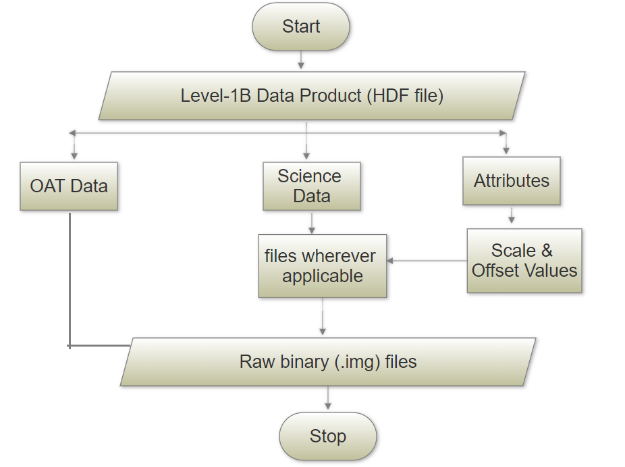


Fig 5 Phase 1 HDF File Extraction

The above shown Fig 5 depicts the flow of working in Phase 1. HDF files provides with three type of data. They are:

1. OAT Data (Orbit and Attitude data)
2. Yaw
3. Pitch
4. Roll
5. Position x, y & z
6. Velocity x, y & z
7. Altitude
8. Science Data
9. Latitude
10. Longitude
11. Sigma0
12. Brightness Temperature
13. Kp
14. SNR
15. X-factor
16. Incidence angle
17. Attributes
18. Scale values
19. Offset values

The Above listed parameters are extracted from the HDF files from which only OAT data are simply stored as .img files. However, respective physical parameters are computed with scale, offset values, and then stored as .img file. These binary files acting as output of phase 1 work as input file for phase 2.

Phase-II Input Data Formation

Binary output files of phase one acts as an input here. After reading the input files Sigma0 flag file parameter is converted to binary form and from which some of the least significant bits are extracted which are used in further phases. Here Sigma0 and Brightness Temperature with their respective good/poor and valid/invalid bits are considered differently. Other physical parameters are kept as it is.

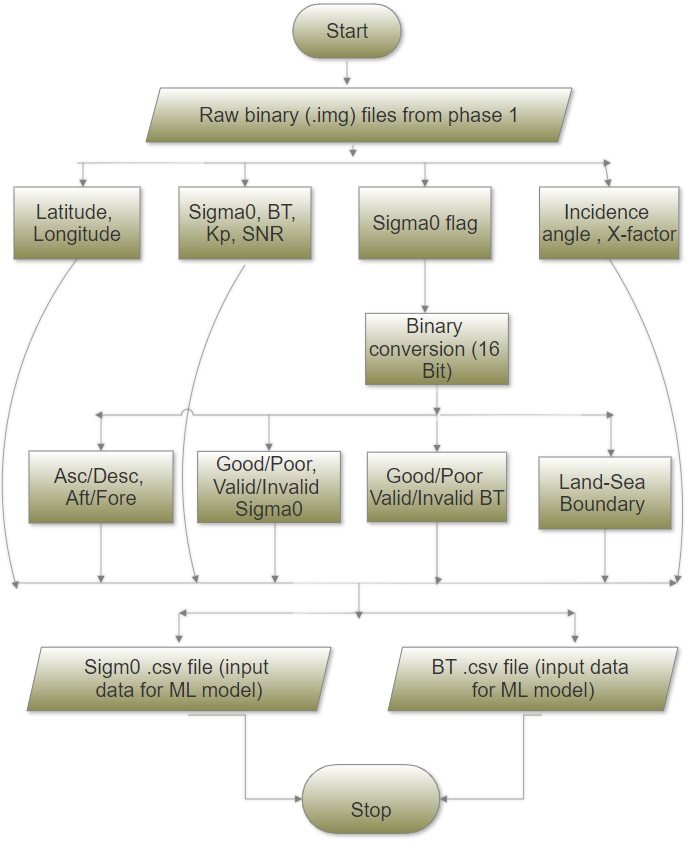


Fig 6 Phase 2 Input Data Formation

The Fig 6 given above shares the flow chart considered for this phase. The bits read from sigma0 flag are:

* [15, 13, 12, 11, 8] for Sigma0
* [15, 13, 10, 9, 8] bits for Brightness Temperature

Other than this bits, other parameters like latitude, longitude, Kp, SNR, X-factor and incidence angle are used from science data. CSV file is formed using all those parameters and passed further for machine learning modelling. In total eight files are formed at the end of phase 2. They are:

* Inner/outer slice sigma0 files
* Inner/outer footprints sigma0
* Inner/outer slice BT
* Inner/outer footprints BT.

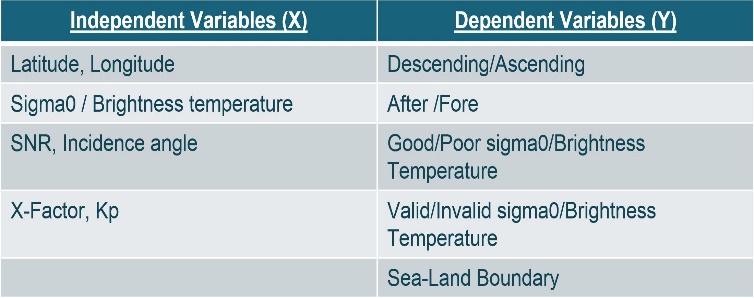
Phase-III Algorithm Selection

Algorithm is the well-stated set of rules for carrying out computational tasks. The Machine learning applications needs two major components, data and algorithm. For the proposed model, data is obtained from phase 2 in CSV file form although for selecting proper algorithm various parameters has to be taken under consideration.

Selection of algorithm is based on type of machine learning technique being used. Selection of technique is dependent on type of dataset and type of output required. The input data obtained at the end of phase 2 is a type a labelled data whereas the proposed models aims at predicting binary class of the flag bits.

For the proposed model, total 12 parameters are considers from which the independent and dependent variables are as shown in following table-1.

Table-1 Independent and Dependent Variables



Machine learning technique used for this model is Supervised Learning as the dataset used is the labelled data.

Phase-IV Machine Learning Modelling

Having all the required information, now we can apply ML techniques. Three scenarios are crosschecked to get most accurate results. They are:

* 80% data is considered as training data and 20% data as test data
* 90% data is considered as training data and 10% data as test data
* Node and Scan-Direction are considered independent variable instead of dependent (A/D – A/F)

For all the above three scenarios, two different procedures were considered. They are

1. Multiple Input Single Output

Here only one dependent variable is predicted at a time. The ML techniques applied here are:

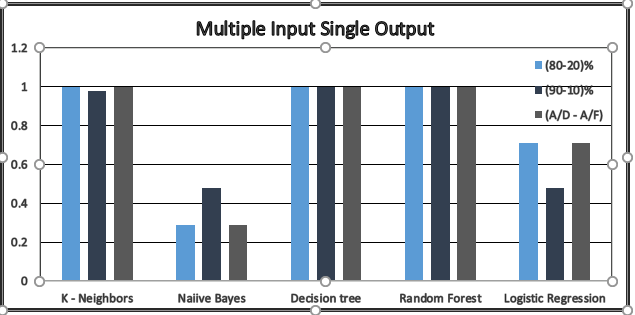
* K-Neighbors
* Decision Tree
* Random Forest
* Naïve Bayes
* Logistic Regression

Fig 7 Multiple Input Single Output

1. Multiple Input Multiple Output

Here all the dependent variables are predicted at a time. The ML techniques applied here are:

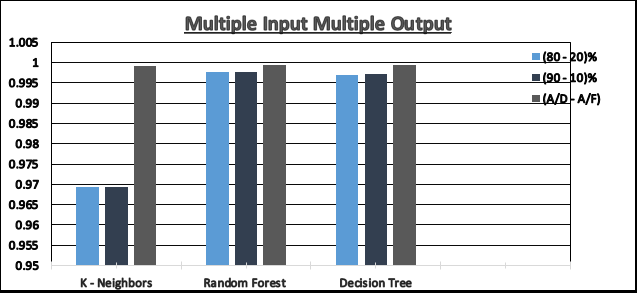
* K-Neighbors
* Decision Tree
* Random Forest

Fig 8 Multiple Input Single Output

V. Result and Analysis

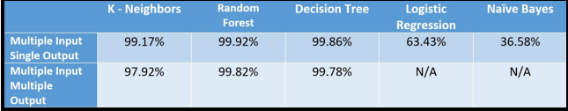
Fig 9 given below, demonstrates the overall accuracy obtained by using each of the algorithms.

Fig 9 Accuracy of algorithm

From the above figure, we can say that Random Forest is the best technique for the given dataset. Based on the accuracy score of techniques the above result is accurate.

VI. References