ISSN XXXX XXXX © 2017 IJESC

IJESC

Research Article Volume 7 Issue No.7

Change Detection Analysis of Coastal Data Based on Big Data Analytics using Satellite Data

Bare Bhakti C.[1], Prof.Dr.S.N.Kini[2]
ME Student[1], Professor[2]
Department of Computer Engineering
JSPM's JSCOE, Hadapsar, Pune, Maharashtra, India

Abstract:
We live in digital world in which data is generated at every second by an activity of users. The size data results into tremendous volume. The data is generated in various formats resulting into variety of types. When data is large and comes in different formats, it is referred to as Big Data. Big data can be a large volume offline data or continuous high streaming data with no control. Only having large data sets is insufficient if we fail to gain information resided within the data sets which leads to knowledge discovery. So there is need of Big Data Analytics which will process data in an effective way to provide knowledge discovery. This paper entitles "Change Detection Analysis Of Coastal Data Based On Big Data Analytics Using Satellite Data". in which analytics of satellite data is provided. In this paper, change detection analysis coastal data is provided using satellite applications. Here, we make use of Tsunami data for the com-parison purpose which helps to provide regression analysis. The system uses an approach of distributed processing which leads to higher performance of the system. It provides efficient way to deal with Big Data. A new technology, Apache Sparks is used to implement the system. This helps to provide change detection analysis to be checked with distributed approach. This system can be extended to provide prediction analytics of tsunami or earthquake. This approach can be used in variety of field to provide Big Data Analytics. In this paper we have discussed Big data analytics, problems to deal with data, an approach to handle data, algorithms used and results implemented.

Index terms: apache sparks, Big Data analytics, change de-tection, distributed processing, remote sensing satellite applica-tions, scala, trained data sets

I. INTRODUCTION

This paper entitles Change Detection Analysis of Coastal Data Based on Big Data Analytics Using Satellite Data. To provide solution to Big Data Analytics, a distributed process-ing approach is used. It provides tool to recognize patterns of coastal data by which information can be gained to achieve knowledge discovery. The problem is to deal with Big Data so that accurate analytics can be done. In this section, we will focus on terms Big data and its challenges, how Big Data Analytics differ from traditional analytics, role of remote sensing applications, solution over the problem.

A. Big Data and its Challenges

Big Data is defined as large volume offline data or high speed continuous streaming data[1].Big Data is composed of not only large of amount of data but also variety of data formats[2].A characteristics of Big Data are[3]-[6]:

1) Volume: There is no specific size of Big Data. Volume of Big data is tremendous which requires distributed system to process it.
2) Velocity: As defined above, Big data is high speed continuous data which becomes difficult to handle in timely manner so that performance of system can be achieved.
3) Variety: A Data is generated at every event done by user. A data can be any format like text, video, audio, structured data or even clicks made by user. Big Data contain multiple formats of data.
4) Veracity: Big Data is collected from various sources with or without maintaining quality of data. Veracity is the reliability of data collected.

5) Variability: The speed of continuous streaming varies from one source to another. Quality of data is also variable.

6)Value: Big Data with given properties is not problem, providing proper analytics by using data is the challenge.
Challenges Of Big Data are[7].
Acquisition of data Analysis
Searching.
Sharing, security.
Querying, updating.
In our system, we focused a main challenge that is searching for appropriate data and its analysis.

B. Big Data Analytics

Traditional data analytics techniques are not sufficient to deal with Big Data. It's not about large data set; it's all about how to get useful data from large data sets in an efficient manner. Big data analytics mainly uses the predictive analytics which specifies what might happen in future.[8] Traditional data analytics focuses on descriptive and diagnostic analysis in which structure of data is known, it uses cleansed data and models used are simple. Big data analytics has the capability of collect data scattered over systems. It helps to provide autonomous decisions. One can provide accurate results by using history data sets to provide predictive analytics.

C. Role of Remote sensing Application

In this system, data is collected from satellite applica-tion. Data are preprocessed by Satellite. One can apply sec-ondary reprocessing to get accurate results. As the preprocessing done accu-rately, results will be more accurate. A remote sensing applica-tion is a software application that processes remote sensing data[9].Remote sensing applications are similar to graphics software, but they enable generating geographic information from satellite and airborne sensor data[9]. Remote sensing application read specialized file formats. Remote Sensing application has the following features:

Change Detection: Determine the changes in object taken at different times of the same area

II. SYSTEM ARCHITECTURE

This section specifies an architecture of system entitled" Change Detection Analysis of Coastal Data Based on Big Data Analytics Using Satellite Data". A proposed system aims to provide analytics of satellite data to check the change detection. It helps in forecasting applications. It aims to Create Visualization. Build mathematical model exploratory analysis and rapid iteration. The basic structural architecture is given in figure 1.It includes data products used, trained data sets, formulae to differentiate normal waves and tsunami waves. It contains various modules. Each module focuses on specific functioning of system.

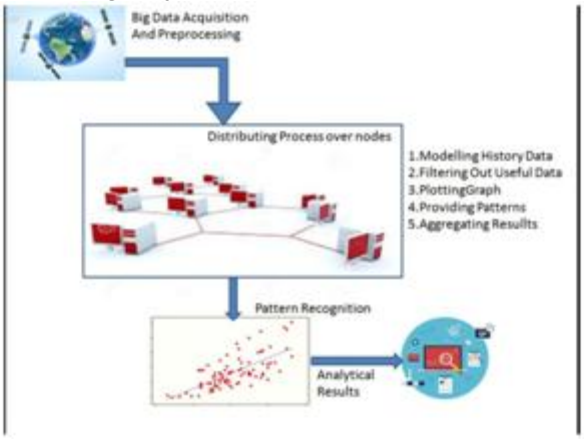

Figure.1. Structural Architecture of System

A. Functions To Be Performed

This system aims to provide change detection so that it will provide a basic way to analyze data. The following functions are considered while implementing the proposed system. Build set of constraints to meet requirements. Evaluate constraints on the use of data. Assess the data structure and its life cycle. Regression analysis is used for predicting change detec-tion and forecasting. Regression analysis helps to understand relationship be-tween independent variables and dependent variables. Time series can be used in pattern recognition.(to rec-ognize patterns of waves in normal condition and in tsunami.). Deep learning will help to find out hidden patterns. Need to deal with variability which is the hidden factor in Big Data Analytics. To deal with data, scheduling of input is required.(time interval for processing).

B. Parameters To Be Considered

The given system aims to provide change detection in sea data which includes following parameters:

Height Of ocean surface: Height of ocean surface is the difference between altitude and range in meters. Significant **wave height:** Height of wave in meter(helps to recognize patterns of wave under some constraints).

Coastal Area: We need to specify selection of survey area for the data acquisition.

Epicenter and Magnitude: In case of underwater earth-quake, we need to consider epicenter of earthquake and magnitude to estimate effects and to study patterns of waves.

Longitude and Latitude: To select survey area.

1) Exceptions: While analyzing patterns of waves, effect of tsunami and forecasting, we need to consider following **constraints:**

Tidal effects of the sun and moon. High tides during rainy season. It should be noted that not every earthquake produces large tsunamis [10]. Detection of tsunami becomes difficult in deep water [11].

C. Physics Of Tsunami

This section describes physics of tsunami[11]–[13].A tsunami is a series of ocean waves with very long wave-lengths (typically hundreds of kilometres) caused by large-scale disturbances of the ocean, such as:earthquakes, landslide, volcanic, eruptions, explosions, meteorites. These disturbances can either be from below (e.g. underwater earthquakes with large vertical displacements, submarine landslides) or from above (e.g. meteorite impacts).Tsunamis can have wavelengths ranging from 10 to 500 km and wave periods of up to an hour. As a result of their long wavelengths, tsunamis act as shallow-water waves. A wave becomes a shallow-water wave when the wavelength is very large compared to the water depth. Shallow-water waves move at a speed, c, that is dependent upon the water depth and is given by the formula:

$$c = \sqrt{gH}$$

where

1) c: Speed of waves in meter/sec or km/hr.
2) g: Acceleration due to gravity ($9{:}8m{=}s^2$).
3) H: Depth of sea.(meter).

suppose

$g=(9{:}8m{=}s^2)$ H=4000m

$$c = \sqrt{9{:}8 \; 4000} = 197{:}98m{=}s$$

It means that tsunami in deep sea(4000 meter) will travel with the speed of 700km/h (approx.).

D. Modelling History Data Sets

A proposed system aims to focus on data of The 2004 Indian Ocean earthquake occurred at 00:58:53 UTC on 26 December with the epicentre off the west coast of Sumatra, Indonesia. The shock had a moment magnitude of 9.1 to 9.3[11]–[14]. History data is used to recognize pattern of sea waves at normal condition and at time of tsunami. This set includes various parameters given below, Longitude and Latitude Significant Wave Height Mean Sea Surface Height Speed of waves etc.

E. Providing Analytics

Each input product is compared with history data set. According to patterns and readings, regression graph is plotted by which one can clearly recognize patterns of sea waves. Input product is collected from Satellite and stored on earth based stations for preprocessing. Filtering useful information is done via Satellite Tools. BRAT 4.0.0 tool[15] is used. These data are made available for further analytics and results interpretation.

III. SYSTEM ANALYSIS

System is implemented to provide analytics using given algorithms. Data Products are taken from European Space

Agency[16]. Preprocessing is done via Satellite Tools BRAT 4.0.0[15].

A. Overview of Product

This section focuses on data sets used in the system which includes history data sets and input product data.

1) History Data: A proposed system aims to focus on data of The 2004 Indian Ocean earthquake occurred at 00:58:53 UTC on 26 December with the epicentre off the west coast of Sumatra, Indonesia. The shock had a moment magnitude of 9.1 to 9.3[14]. History data is used to recognize pattern of sea waves at normal condition and at time of tsunami. This set includes various parameters given below, Longitude and Latitude Significant Wave Height Mean Sea Surface Height Speed of waves etc.

B. Data Product Used

Data product is provided by European Space Agency's Cryosat-2 Ocean Products. Format of product is given below.[17][18]

1) Definitions: Two instrument data levels are considered for Cryosat-2 Ocean Products:

1) Level 1b products correspond to geo-located engineering-calibrated products.

2) Level 2 products correspond to geo-located geophysical products.

Two delivery delays are considered for Cryosat-2 Ocean Products:

1) Intermediate (IOP): delivered within 48 hours after data sensing acquisition, due mainly to the consolidation of some auxiliary or ancillary data (e.g. preliminary restituted orbit data)

2) Geophysical (GOP): delivered within typically 1 month after data acquisition, due mainly to the consolidation of some auxiliary or ancillary data (e.g. precise orbit data).

Figure 2 shows the file structure of Cryosat Ocean Product.

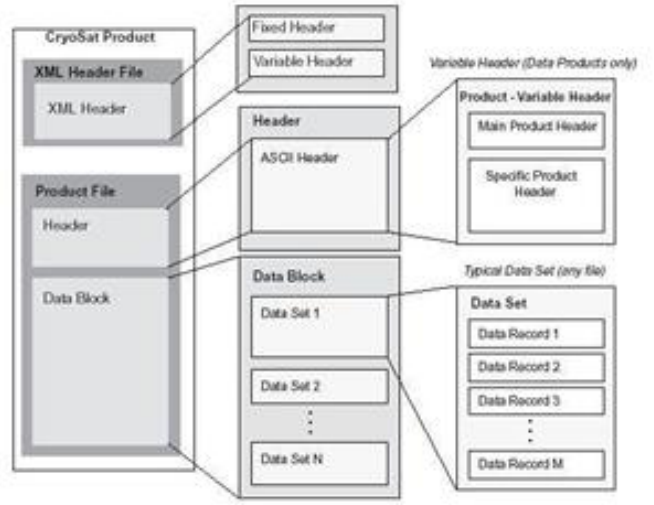

Figure. 2. Earth Explorer Ground Segment File Format Standard[18]

2) Product Files: As shown in Figure 3, each product file is composed of:

A Main Product Header (MPH) A Specific Product Header (SPH) Data Sets the MPH and SPH blocks are ASCII, whereas the Data Sets are completely binary and contain one or more Data Set Records each. Level 1b and Level 2 Ocean Products contain one Data Set, which leads to the product structure shown in the Figure 4 below:

C. Implementation Tools

Proposed system deal with offline stored big data to pro-vide Big Data Analytics with training data set and new datasets. It makes use of iterative processing with deep learning

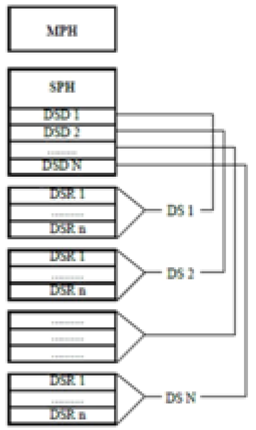

Figure.3. Generalised Product Structure[18]

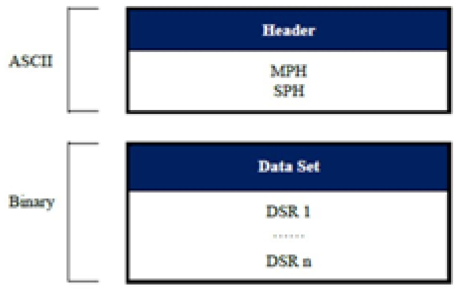

Figure.4. Ocean Products Structure[18]

As data are taken with specified time series. Distributing all processes results into higher performance of the system and code optimization provides time efficiency. To achieve greater performance of the system, Apache Sparks is used as described below.

D. Apache Sparks:

It is a popular open-source platform for large scale data processing that is well-suited for iterative machine learning tasks. By allowing user programs to load data into a cluster's memory and query it repeatedly, Spark is well-suited to machine learning algorithms. Apache Sparks version 1.5.2 is used.

E. Scala:

Scala is a general-purpose programming language providing support for functional programming and a strong static type system. Designed to be concise, many of Scala's design decisions aimed to address criticisms of Java. Scala 2.10.4 version is used to implement all modules.

IV. ALGORITHMS

This sections defines algorithms to be implemented modules of system structure.

A. Module 1:Data Collection And Preprocessing

Input: Trained Data Set Collected From Satellite. Output: Regression Analysis by plotting graphs

Steps:

1) Collect Data with specified parameters wave height, longitude latitude, ocean depth.

2) Calculate speed of waves by using formula

$C=\sqrt{gH}$

3) Create (x,y) series from data set for chart.

4) Plot chart with custom 'x' range.

5) Plot graph with fitted line for the regression analysis with parameters speed and wave height

6) Store result in ASCII format

B. Module 2: Data Processing

Input: Input Data Product from Cryosat

Output: Regression Analysis by plotting graphs

Steps:

1) Collect Data with specified parameters wave height, longitude latitude, ocean depth by selecting survey area.

2) Calculate speed of waves by using formula

$C=\sqrt{gH}$

3) Compare results with history data set.

4) Plot graph with fitted line for the regression analysis with parameters speed and wave height.

5) Store result in ASCII format.

C. Module 3:Result Storage and Interpretation

Input: History Data Set and Processed Data Set Output: Results in ASCII Format

Steps:

1) Store history data set and processed data set.

2) Store results by comparing both data sets.

3) Pattern Recognition

D. Module 4:Data Analytics

Input: Recognized Patterns Output: Forecasting by change detection.

Steps:

1) This module aims to provide analytical results by recognizing patterns.

2) By these patterns, one can see for the change detection and patterns of sea waves.

3) This information can be used by forecasting application or other applications

V. RESULTS AND IMPLEMENTATION

The proposed system focuses on functional testing. Each module is tested with data. Big Data analytics uses the concept of Proof of Concept (POC).Testing is done with different approaches to meet performance of the system with time efficiency and accuracy. Table I and II shows the different testing strategies to meet high performance of system by time efficiency.

A. Local Machine:

A proposed system is executed on local machine having configuration of

Processor: Intel(R)Core(TM)i5–421U CPU @ 1.70GHz 2.40 GHz

RAM: 8.00 GB

System Type: 64-Bit OS,x64–based processor

B. HPC Cluster Node[19] :

A proposed system is executed on HPC Cluster node provided by Centre for Development of Advanced Comput-ing (C-DAC), Pune. PARAM YuvaII is a High Performance Computing (HPC) Cluster that is among the latest addition to the series of prestigious PARAM series of Supercomputers built in India.

Table. I. Testing products on single core on local machine

Products	Product Size(mb)	Size Of Processed File(kb)	Execution Time(s)	Size Of Output File(kb
1	14	3	19.35	4
1,2	28.3	6	16.38	8
1,2,3	42.6	16	18.37	28

Table. Ii. Testing products on multicore on local machine

Products	Product Size(mb)	Size Of Processed File(kb)	Execution Time(s)	Size Of Output File(kb
1	14	3	17	4
1,2	28.3	6	18.45	8
1,2,3	42.6	16	19.21	28

Table iii testing products on cluster node provided by Cdac[20],pune(internet access with 72 mbps)

Products	Product Size(mb)	Size Of Processed File(kb)	Execution Time(s)	Size Of Output File(kb
1	14	3.5	12	4
1 to 9	126	60.2	20.51	127

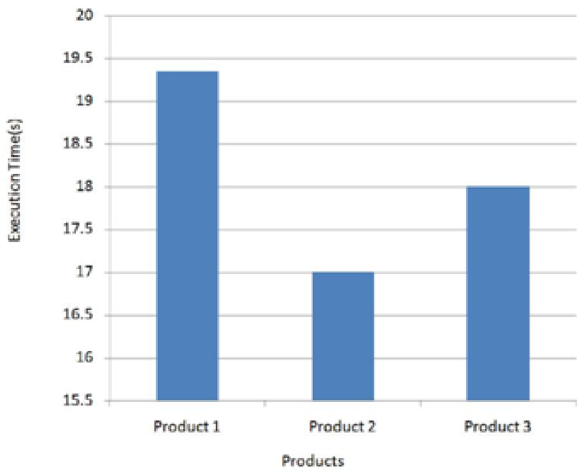

Figure. 5. Products running on single core

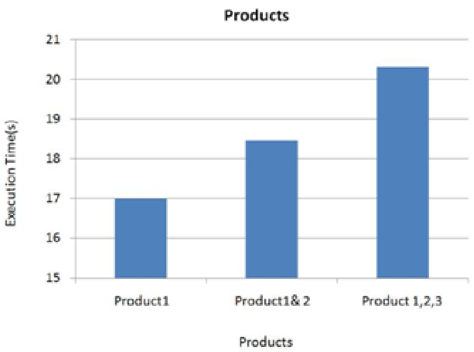

Figure. 6. Products running on multicore

Figure 5, 6 and 7 shows graph implementation with different approaches of running system.

VI. CONCLUSION

In this paper, we focused mainly on the analytics of big data to provide change detection analysis of coastal data collected from satellite application .It is difficult to work with large and complex data sets. Traditional analytics shows their insufficiency to deal with large amount data where the data size varies from TB to PB. While working with real time remote sensing big data, we need to provide advancements in distributed processing. To provide analytics of Big Data is challenging task. Handling large volume of data is critical task. As we are interested in knowledge gained by information, mining of data from large volume data sets has limitations of designing filters in such way that it must not discard useful data from data set. So there is need to design filters in such way, so that it will include all required data without loss of useful data. Another problem is to provide high performance of systems. It can be achieved by distributing task over clusters of node so that work is done within less time with efficiency.

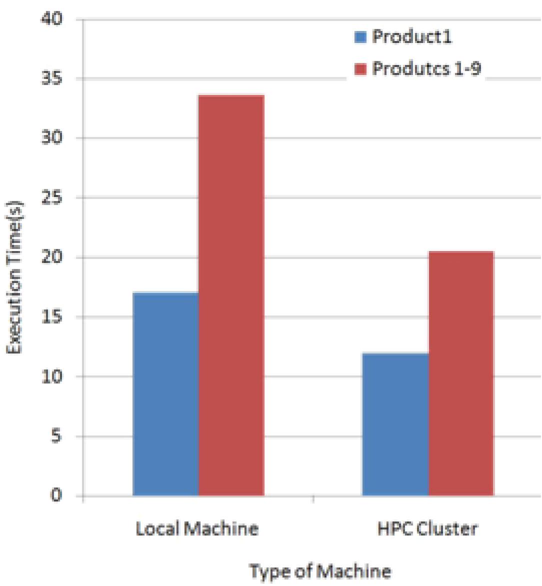

Figure.7. Products running on Local Machine and HPC Cluster Node

More advanced algorithms are needed to extract corre-lations from the data while allowing different levels of privacy. Filtering out unwanted data is another significant chal-lenge. One challenge is to define these filters in such a way that they do not discard useful information. Another challenge is to automatically generate the right meta data, to describe what data is recorded and how it is recorded and measured. Data analysis is considerably more challenging than sim-ply locating, identifying, understanding, and citing data.

VII. FUTURE ENHANCEMENT

The proposed system entitles Change Detection Analysis Of Coastal Data Based On Big Data Analytics Using Satellite Data is based on forecasting analytics which specifies to know what might happen in future. This type of analytics is used in forecasting applications, tsunami prediction, earthquake prediction, fire detection (if image processing used).It aims to manage large amounts of data despite of hardware, software, and bandwidth constraints. Real time processing of remote sensing big data can be used in various fields like Health care systems, agriculture all types of mining etc. Distributed processing using real time satellite data has so many chal-lenges. There is no recent technology that can be used to implement it. We need to provide more advanced system to deal with as the collaboration and communication in distributed system is challenging. The proposed system can be extended to provide predictive analytics of Tsunami, earth quake etc. But prediction of earthquake is inherently impossible fact[21].So this system can be enhanced to study patterns of coastal data and for the forecasting conditions of coastal data

VIII. ACKNOWLEDGMENT

This is to acknowledge and thank all the individuals who played defining role in shaping this seminar report. Without their constant support, guidance and assistance this seminar report would not have been completed. Without their co-ordination, guidance and reviewing this task could not be completed alone. I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to my guide Prof. S.N.Kini for giving his valuable guidance, inspiration and encouragement to embark this seminar. This seminar being conceptual one needed a lot of support from my guide so that I could achieve what I was set out to get. I would personally like to thank, Prof. M. D. Ingle, PG Co-ordinator, Computer Department, Prof. H. A. Hingoliwala, Head Of Computer Department and our Honorable Principal Dr. M. G. Jadhav Sir who creates a healthy environment for all of us to learn in best possible way. I would like to thank to Mr.Y.Satya Swarup Sir and Mr.Rishi Pathak Sir (Centre for Development of Advanced Computing (C-DAC), Pune) for giving valuable guidance about HPCs. I would like to express my gratitude to all my colleagues for their support, co-operation and fruitful discussions on diverse topics and technical help. I would especially like to thank my family and friends for their sincere interest in my study and their moral support.

IX. REFERENCES

[1].https://en.wikipedia.org/wiki/Big data, “Big Data”

[2].Katina Michael, University of Wollongong Keith W. Miller, University of Missouri–St. Louis, “Big Data: New Opportunities and New Challenges,” The IEEE Computer Society 0018 –9162/13/$31.00 2013 IEEE pp 22–24

[3].J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, “Mad skills: New analysis practices for Big Data,” PVLDB, vol. 2, no. 2, pp. 1481–1492, 2009.

[4].Muhammad Mazhar Ullah Rathore, Anand Paul, Awais Ahmad, Bo-Wei Chen, Member, Bormin Huang, and Wen Ji,

"Real Time Big Data Ana-lytical Architecture for Remote Sensing Application, "IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 8, NO. 10, OCTOBER 2015

[5].Muhammad Habib ur Rehman,Chee Sun Liew,Assad Abbas, Prem Prakash Jayaraman,Teh Ying Wah,Samee U. Kha ,"Big Data Reduction Methods: A Survey "Data Sci. Eng. DOI 10.1007/s41019-016-0022-0 Springer

[6].XUE-WEN CHEN, (Senior Member, IEEE), AND XIAOTONG LIN, "Big Data Deep Learning: Challenges and Perspectives, " Digital Object Identifier 10.1109/ ACCESS. 2014. 2325029 IEEE Access Volume 2 pp 514-525

[7].http://www.computerweekly.com/feature/Big-data-and-analytics-a-large-challenge-offering-great-opportunities, "Big data and analytics: a large challenge offering great opportunities"

[8].http://www.predictiveanalyticstoday.com/category/data/ "Predictive Ana-lytics "

[9].https://en.wikipedia.org/wiki/Remote sensing application, "Remote sens-ing applications"

[10].http://itic.ioc–unesco.org "How do earthquakes generate tsunami. "

[11].http://www.bom.gov.au/tsunami/info/"Tsunami Facts and Information"

[12].http://nctr.pmel.noaa.gov/tsunami-forecast.html "Tsunami Forecasting"

[13].http://ptwc.weather.gov/faq.php19 "PTWC Frequently Asked Questions (FAQ)"

[14].https://earth.esa.int "Earthquake and tsunami (Asia) - 26 December 2004"