Generation of Analysis Ready Data for Indian Resourcesat Sensors and its Implementation in Cloud Platform

Thara Nair¹, Akshay Singh², E. Venkateswarlu¹, G.P. Swamy¹, Vinod M Bothale¹ and B. Gopala Krishna¹

¹National Remote Sensing Centre (NRSC), Indian Space Research Organisation (ISRO), India
²Jaypee Institute of Information Technology, India

Email: thara_nair, venkateswarlu_e, swamy_gp, vinod_bothale, bgk}@nrsc.gov.in

Received: 06 March 2019; Accepted: 22 May 2019; Published: 08 June 2019

Abstract—The introduction of remote sensing techniques had lead us into a new race of advanced data processing applications. The analysis ready data is also a part of it which is generated at the producer end to facilitate its user to directly go on to the application part. This paper highlights the generation, processing and cloud applications of the Analysis Ready Data (ARD) using ISRO's Satellites ResourceSat-2 and ResourceSat-2A's LISS-3 sensor data. The proposed work includes use of terrain corrected data for generating Radiance, Top of Atmosphere (ToA) Reflectance, Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and Time series analysis with pixel level Quality Assessment (QA) for all the generated data products. A graphical user interface has been developed for online ordering of data by the user. This paper also highlights the implementation of the developed application in cloud platform using the cloud computing model, Platform as a Service (PaaS). This facilitates the users to generate the ARD products from any device, facilitating a quick and all time available transmission rate for the customers.

Index Terms—Analysis ready data, image processing, PaaS cloud platform, reflectance, remote sensing, ressourcesat, satellite imagery, vegetation index.

I. INTRODUCTION

The influence of remote sensing on human lives has seen a magnificent increase in the past decade. From weather predictions to finding the change in habitation of a region over time, it finds application in all walks of our daily life. According to a recent study by the Goddard Space Flight Center, there are approximately 2271 satellites in orbit. This scenario demands the players to be continuously maintaining high standards and redefining themselves from time to time. In this regard, the Indian Remote Sensing (IRS) program of Earth observation has witnessed quantum jumps since its inception in 1988, facilitating India’s collection of large scale multi-spectral data around the globe for various kinds of cartographic, disaster management and resource planning activities.

With the challenging developments in space technology, space is an equally competitive platform as land and water. The massive number of satellites being launched worldwide has opened up a new avenue in investigating earth and other planets of the solar system. Data transmitted from these satellites has seen an exponential increase in terms of data rate and volume. This has necessitated significant advances in image processing in terms of advanced algorithms, improved processing techniques and data storage.

Analysis Ready Data (ARD) is the new paradigm, which can be the new workhorse for application scientists. It is often observed as an obligatory need for application scientists to replicate similar processing steps for the same data, time and again, resulting in huge wastage of valuable scientific resources. The Committee on Earth Observation Satellites (CEOS) define ARD as “satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis without additional user effort and interoperability with other datasets both through time and space”[1,2,3]. This development will ensure an improved access to satellite data and in turn will open the door to a wider audience to exploit it, in addition to servicing existing users with better efficiency and at reduced cost.

The immediate solution offered by technology for all the above stated challenges is Cloud Computing. To harness the flexibility, efficiency and strategic value of Cloud Computing platform, satellite data storage and processing can be envisaged in the cloud platform. In addition to the concept of data storage on cloud, we also need to plan for the entire data processing chain over a cloud platform. This ensures remote access to the huge volume of data and on-demand generation and dissemination of all the required products for analysis or application ready use. The Platform as a Service (PaaS) is a platform based service which allows its users to build,
run and deploy its applications without worrying of infrastructure management. The use of this platform will allow us more versatility, cognizance and edge over all the present competitors in the global market. Since cloud computing allows access to all required applications through the internet, it offers very high flexibility in data dissemination to the users. This not only provides a faster and reliable producer side support to the customers but also, helps them build stronger relationships. This paper highlights the design methodology of ARD implemented for Indian Remote Sensing Satellites and its implementation in Cloud Computing Platform.

II. RELATED WORK

The United States Geological Survey (USGS) has processed all of the Landsat 4 and 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) archive over the conterminous United States (CONUS), Alaska, and Hawaii, into Landsat ARD. Provision of pre-prepared ARD is intended to make it easier for users to produce Landsat-based maps of land cover and land-cover change and other derived geophysical and biophysical products. The ARD are provided as tiled, georegistered, top of atmosphere and atmospherically corrected products defined in a common equal area projection, accompanied by spatially explicit quality assessment information, and appropriate metadata to enable further processing while retaining traceability of data provenance[4]. The Planet Lab is dedicating a lot of resources and thinking about how to make it easier for user community to bypass pre-processing steps and dive to analysis[5]. Committee on Earth Observation Satellites (CEOS) is developing a Data cube architecture that depends on CEOS analysis Ready Data for Land (CARD4L) to allow immediate analysis and improved interoperability among datasets. Since 1988, NRSC has acquired large volume of multispectral data from IRS Satellites and propose to generate ARD products for user dissemination.

III. METHODOLOGY

Indian Remote Sensing Satellites Resourcesat-2/2A are having a Linear Imaging Self-Scanning Sensor (LISS-III) with a spatial resolution of 23.5 meters providing four spectral bands i.e. Band 2 (green), Band 3 (red), Band 4 (near infrared) and Band 5 (short wave infrared)[6]. The raw data acquired from the satellite undergoes both radiometric and geometric corrections to generate terrain corrected products, ready to be delivered to the user community.

3.1 ARD Products Generation

This section describes the methodology adopted for generation of Analysis Ready Data (ARD) products. Fig.1 depicts the flow of ARD products generation. In this activity, the focus is on the generation of Analysis Ready Data products for any scene for Resourcesat 2/2A, based on demand. Any ARD product is designed to have scene wise terrain corrected product, Top of Atmosphere Reflectance (ToA) for each band, Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and a cloud mask band.

The generation of Top of Atmosphere Reflectance starts with the computation of the radiance values of all the bands. The consequent step in the generation of ToA product is the generation of cloud mask[7] and its application on ToA reflectance bands. The product may or may not have a number of cloudy/cloud shadowed pixels, which have to be masked to provide valid product to the user. The masking value is chosen as 255 for cloud and cloud shadow regions.

The ARD product also comprises of NDVI and EVI products of the corresponding scene. NDVI and EVI are computed from the Surface Reflectance of the corresponding bands. This is achieved by subjecting the radiance values to DOS (Dark Object Subtraction)[8]. After the DOS correction, the resultant image would be atmospheric haze corrected which is then applied for calculation of other ARD products. In this process BAND4 (NIR band) and BAND3 (Red band) are considered as inputs to generate NDVI and EVI. Cloud mask is applied to the resultant image so as to get rid off all the cloudy pixels and generate a valid product to the user.

3.1.1 Generation of Radiance Values

For generating ToA reflectance, radiance values are to be estimated, which is a direct measure of radiance reflected from the surface, neighboring pixels and clouds depending both on illumination as well as the position and orientation of the target. The formula (1) for getting radiance of a single pixel is:

\[ L_\lambda = \frac{(L_{\max} - L_{\min})}{(Q_{calmax} - Q_{calmin})} \times (Q_{cal} - Q_{calmin}) + L_{\min}. \]  

Where,

\[ L_\lambda \] the cell value of radiance
\[ Q_{cal} \] = digital number (or pixel value)
\[ L_{\min} \] = spectral radiance scales to Qcalmin
\[ L_{\max} \] = spectral radiance scales to Qcalmax
\[ Q_{calmin} \] = the min quantized calibrated pixel value (= 0 in our case)
\[ Q_{calmax} \] = the max quantized calibrated pixel value (= 1023 in our case)
3.1.2 Generation of Top of Atmosphere (ToA) Reflectance

Top of Atmosphere (ToA) Reflectance provides us the ratio of radiance reflected to the solar radiation incident on a given surface and is a unit less quantity. For computing ToA reflectance, radiance to ToA conversion [9] is applied to evaluate the ToA of the corresponding spectral band:

\[
\rho_\lambda = \frac{(\pi \cdot L_\lambda \cdot d^2)}{(E_{\text{SUN}} \lambda \cdot \cos \theta_s)}
\]  

(2)

Where,

- \( \rho_\lambda \) = Unit-less planetary reflectance
- \( L_\lambda \) = spectral radiance (from the earlier step)
- \( d \) = distance b/w Earth-Sun in astronomical units
- \( E_{\text{SUN}} \lambda \) = mean solar exo-atmospheric irradiances
- \( \theta_s \) = solar zenith angle

Furthermore, the formula for calculating \( d \):

\[
d = (1 - 0.01672 \cdot \cos (\text{radians} (0.9856 \cdot (\text{Julian Day} - 4))))
\]

Where, Julian day = Total integer no. of days elapsed
3.1.3 Generation of Cloud Mask

Clouds play an important role in the climate system having importance in energy balance of the atmosphere. The data acquired by optical remote sensors is affected by the presence of clouds. In such scenarios, it is very essential to provide information on cloudy pixels for proper and efficient utilization of data. Exact evaluation of cloudy and cloud shadowed pixels is an important pre-processing step for any reliable cloud removal or masking algorithm[7]. Due to the absence of thermal band for Resourcesat-2, an algorithm based on unsupervised classification exploiting spectral attributes is developed by NRSC for automatic cloudy and cloud shadowed pixels detection. Some of the salient features of this utility are – it generates cloud and shadow mask file for each input data, it can delineate snow and moisture cloud, salt pans and moisture cloud which have same spectral signature.

3.1.4 Generation of Normalized Difference Vegetation Index

Normalized Difference Vegetation Index (NDVI) is a measure to help us observing the vegetation index of the Earth's surface. It is beneficial in assessing the vegetative cover by employing time series analysis. Its values generally range from -1 to 1. Negative values of NDVI (values approaching -1) correspond to water. Values near zero (-0.1 to 0.1) typically correspond to barren areas of rock, sand, or snow. Lastly, low, positive values represent shrub and grassland (approximately 0.2 to 0.4), while high values indicate temperate and tropical rainforests [8]. The formula for calculating NDVI is expressed below:

\[ NDVI = \frac{(\text{NIR}-\text{Red})}{(\text{NIR}+\text{Red})} \]  

Where,

\[ \text{NIR} = \text{Band}4 \text{ (near infrared band)} \]
\[ \text{Red} = \text{Band}3 \text{ (red band)} \]

3.1.5 Generation of Enhanced Vegetation Index

Enhanced Vegetation Index (EVI) can simply be termed as optimal NDVI , where the variations in solar incidence angle, distortions in reflected light caused due to particles of air, canopy background signals and atmospheric conditions are corrected . While, NDVI can show varying results within a time difference of mere 2 hours due to changes in solar incidence angle, shadows and soil boundaries. EVI on the other hand, provides better consistency. The formula for calculating EVI (when blue band is not available in the sensor) is expressed below [9] (Jiang et al. 2010)

\[ EVI = 2.5 \times \frac{(\text{NIR}-\text{Red})}{(\text{NIR}+2.4\text{Red}+1)} \]  

3.1.6 Generation of Surface Reflectance Product

The Dark Object Subtraction (DOS) method is an image-based technique to cancel out the haze component caused by additive scattering from remote sensing data (Chavez Jr, 1988). While calculating SR, we first get Lmin as the minimum radiance pixel value. From Lmin, Lpath is calculated as shown below in equation (5) which is then subtracted from all the radiance pixels for DOS correction. After then, SR values of every individual pixel is found by applying the formula shown in equation (6):

\[ L_{\text{path}} = L_{\text{min}} - L \]  
\[ \rho = \frac{(L_{\text{sensor}}-L_{\text{path}})d^2}{E_0\cos q_2 T_z} \]  

Where,

\[ \rho = \text{Unit-less surface reflectance} \]
\[ L_{\text{sensor}} = \text{is at-sensor radiance} \]
\[ d = \text{distance b/w Earth-Sun in astronomical units} \]
\[ E_0 = \text{mean solar exo-atmospheric irradiances} \]
\[ B2=184.0; \; B3=155.1; \; B4=104.4; \; B5=22.57 \]
\[ \theta z = \text{solar zenith angle (= 90 - elevation angle)} \]
\[ T_z = \text{approximated by } 0\theta \]

Further, the formula for calculating d:

\[ d = (1-0.01672*\cos (\text{radians (0.9856*(Julian Day-4))))} \]  

Where,

\[ \text{Julian day} = \text{Total integer no. of days elapsed} \]

3.2 Ard Process Flow

For ease of product ordering, a Graphical User Interface (GUI) was developed. This section highlights the flow of processes starting from the ordering of data by the user via GUI ordering platform and the subsequent producer's side of receiving the request, processing the raw inputs and sending the resulting ARD data products to the user. Whenever the user enters data request either for a single day data or for a time series of data, a request submitted mail is sent to the user and at the same time, requested data products mail is received by the producer with the link of the data folder stored on the cloud data storage (here we are using Google cloud storage). The producer is then required to upload the same input data folder to the PaaS platform where it will then be able to process those data products to get the resulting ARD products as per requested by the user. It shall be noted that all the programs are also uploaded to the PaaS platform before doing any data generation activity. The whole process is similar to the way the products might have been generated over a PC having python or any IDE for python installed (or accordingly for any other programming language used).
All the ARD products can be ordered by user through the GUI ordering platform for their specific requirements in terms of region, season and spectral bands. The Fig.2 highlights the flow of processes starting from the ordering of data by the user via GUI ordering platform and the subsequent producer's side of receiving the request, processing the raw inputs and sending the resulting ARD data products to the user. The data products generated over the PaaS platform gets stored there and are available to be downloaded by the producer who can then send those to the user either by mail, FTP or any other means according to the request made. There are a number of PaaS platforms available at the market which can be selected according to the languages they support and user interface. As we are using Python as our main language, we selected Python any where PaaS cloud platform which have a great user interface and customer support.

IV. GUI ORDERING PROCEDURE

The GUI for User Order Processing System(UOPS) allow the users to order ARD data. As soon as the user makes a request (fig.3), a confirmation message pops up to the user with an email notification. The request includes the user's name, email address, date of the data required, corresponding ARD data products required and the satellite name of whose data is needed.

The GUI platform also gives option to the user for ordering a time series data. The time series data form asks for name, email address, start and end dates, ARD products to be generated and corresponding satellites for data. Here, we have two additional options allowing the user for deciding a time series on basis of either lat-long or path-row. Where lat-long corresponds to the nearest scene center latitude and longitude coordinates of the data to be acquired, the path-row corresponds to the satellite's path and row information. For a lat-long selection, an input is picked only when the user inputted lat-long is less than 0.05 degrees of the scene center lat-long. On the other side if a path-row selection is made, an input is picked only when the user inputted path and row matches exactly with the path-row of the data within the date range entered.

The request made by the user is also sent to the producer on email with the name, email, start and end date (single date for single day data), ARD products asked, lat long/path-row values (for time series data), satellite name of whose data is asked and the input folder links.

Here the input folder links represents the cloud links of the data folders. For e.g. we are using Google cloud here hence, Google cloud link is been generated. As soon as the link is clicked the producer is guided to the folder containing the input data to be used for the generation of the ARD products. These data products needs to be downloaded by the producer to implement them on PaaS cloud platform as shown in the next section.

Fig.2. Data ordering and sending process flowchart

Fig.3. Single data request form
4.1 Cloud Platform Data Generation

In our research we are using Python any where cloud platform which is a PaaS cloud platform for generating and storing data products as per requested by the user. A PaaS cloud platform provide Platform as a Service allowing users to deploy, run and manage applications over it without complexity of building or maintaining the infrastructure associated with the same. As continued from the earlier step of downloading the required input data from the cloud storage we are now required to upload the same over the PaaS cloud platform as our input data. This data is then subjected to the programs which generates the corresponding ARD products as per asked by the user. The output files then gets saved to the PaaS platform which are available to be downloaded by the producer who can then send them to the user via email, FTP or any other way prescribed.

PaaS platform dashboard show various options like number of consoles available (here one can run multiple programs at single time) and the versions. Also shows the uploaded files, programs, images and output files in a sample directory. All these can be easily added, deleted, modified or downloaded. Also, any additional packages to be used shall be added here only.

Fig.4 shows a sample screen where a code is modified with options of saving and running in respective language versions. Pythonanywhere also gives us an option of scheduling our tasks on regular time intervals. For e.g. if one wants to generate NDVI product every 5th day of a month for a specific region, it can schedule that task for once rather doing the same manual work every time.

V. RESULTS

5.1 Generation of Top of Atmosphere Reflectance

Top of Atmosphere (ToA) Reflectance provides us the ratio of radiance reflected to the solar radiation incident on a given surface and is a unit less quantity.

Fig.5. LISS-3 Band 2 and its corresponding Radiance image

Fig.6. LISS-3 Band 2 image, Radiance,ToA (left to right)

Here, Fig.5. shows the radiance image calculated from the corresponding Band 2 DN image and Fig.6 shows the ToA calculated for the same layer.

5.2 Generation of Normalized Difference Vegetation Index

Normalized Difference Vegetation Index (NDVI) is a measure to help us observing the vegetation index of the Earth's surface.

The Fig.7 shows the Band 4 (NIR) and Band 3 (Red) which are used to get the resultant NDVI (lower left) which is then corrected with cloud masking. Fig.8 shows the color composite NDVI image in which the green portions depict dense green forests, bright yellow depict shrubs or less dense trees, golden yellow depict grass and magenta or dark blue depict absence of vegetation. Fig.9 shows the level of precision achieved when we use another image for testing which is nearly cloud free. Here it can be observed that pixel level tracing and correction of cloudy pixels is achieved.

Fig.4. Sample Coding Environment
5.3 Generation of Enhanced Vegetation Index

In Fig.10, the generated EVI is shown using Band4 and Band3 and then cloud masked for cloudy/cloud shadowed pixels. While Fig.11 shows the color composite of the generated EVI image with color strip corresponding to EVI values of the region. The Fig.12 shows a comparison between NDVI and EVI for the same area and time. While the following Table-1 and 2 show the range of NDVI and EVI values respectively. It shows the range of NDVI varying from a minimum of -0.19039 to a maximum of 0.63362. Similarly, for the same region and time the values of EVI ranges from a minimum of -0.25353 to a maximum of 1.2615 (depicting the reason why it is called enhanced vegetation index).
5.4 Generation of Time Series Data

Time series data analysis provides the advantage of fully capturing the dynamic changes of any pattern viz. vegetation, land cover etc. over a period of time. This helps us noticing the changes in the vegetation index through a course of time varying either due to natural (season changes, natural disaster, etc) or artificial activities (constructions, farming, etc). This data is then implemented for efficient resource planning, strategic and military purposes.

Fig.13 shows a time series NDVI data generated for different dates showing the change in vegetation index over a period of time (approx. 3 years). It is ResourceSat-2 LISS-3 sensor data with Path-103 and Row-060.

VI. CONCLUSIONS & FUTURE WORK

We demonstrated the generation of Analysis Ready Data products generation from ResourceSat sensors in this work. The ARD products are accompanied with meta data to ensure processing traceability and queries. Analysis Ready Data (ARD) products has always proved to be a boon to remote sensing satellite data users. It not only facilitates direct usage of data without any additional user effort but also, helps saving a lot of time at customer end. The major benefit of ARD approach is the scalability it offers through adding of new sensors allowing users to focus on EO data analytics. The cloud platform helps producers as well as users for faster data attainability, sharing and research implementations. It allows producers to remotely store, generate and access data and dismiss the dependence over fixed systems. Future work includes creation of consistent time series data, scalability with new sensors, developing inter calibration procedures, data analytics and addition of high security features for cloud data maintenance.

ACKNOWLEDGEMENTS

The authors of this work would like to express their gratitude to Director, NRSC for all his guidance. The authors deeply acknowledge the constant support provided by Sri. T.Sivannarayana, Manager, C&DA and the staff of IODP Group and other reviewers. The authors would also like to express their token of thanks to the Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida for their efforts.

REFERENCES

[5] Scott Soene, “ Planet’s framework for Analysis Ready Data”.
Authors’ Profiles

Akshay Singh, M.Tech. in Electronics & Communication Engineering from JIIT University and worked as project student in DPPA&WAA, NRSC. His research interests include system automation, E-vehicles and smart city programs.

E.Venkateswarlu, M.Sc., M.Phil. in Mathematics, is currently working as Scientist “SF” in DPPA&WA area of NRSC, ISRO. He has published 6 research papers in reputed conferences and journals. His research interests include Image Processing, Data Quality Assurance, etc.

Thara Nair, M.Tech in Control Systems is working as Scientist “SG” and Head, IRS Optical Data Processing Division in DPPA&WAA, NRSC. She has published few research papers in reputed conferences and journals. Her research interests include Image Processing, system design, etc.

G.P. Swamy, M.Tech in Computer Science is working as Scientist “SG” and Group Head, IRS Optical Data Processing Group, DPPA&WAA, NRSC. He has published few research papers in reputed conferences and journals. His research interests include Image Processing, system design, Data Quality Assurance, etc.

Vinod M Bothale, a gold medalist from IIT Roorkee and a fellow of Institution of Engineers, is Deputy Director of Data Processing and Web Applications Area at National Remote Sensing Centre, ISRO. His research interests include image processing, geospatial applications to natural resources management and WebGIS software applications. He has been the designer and architect of Spatial Data Infrastructures (SDI) and Web Portals. He has been founder designer of Bhuvan Geoportal and water resources portal, India-WRIS. He has number of awards to his credit.

B.Gopala Krishna, received M.Sc.Tech (Electronics) from Andhra University, Visakhapatnam, and M.Tech (Electronics) from IIT, Kharagpur with a specialization in Satellite Communications and Remote Sensing. He is Deputy Director (Retd.) of DPPA&WAA at NRSC, ISRO, Hyderabad. He has more than 170 publications to his credit in National and International journals. He is a life member of various technical societies viz. INCA, ISRS, IMSA, ISG, ISSE and ASCI. His research interests include digital photogrammetry and mapping, geometrical data processing for remotely sensed data, planetary data processing, stereo image analysis.