RADIO OCCULTATION WITH ROSA ON-BOARD OCEANSAT-2 IN GRID ENVIRONMENT

L. Mossucca(1), O. Terzo(1), R. Notarpietro(2), M. Cucca(2)

(1) Istituto Superiore Mario Boella, Torino, Italy ((mossucca, terzo) @ismb.it),(2) Politecnico di Torino, Torino, Italy ((riccardo.notarpietro, manuela.cucca)@polito.it)

Abstract: In September 2009 the new Italian GPS receiver for Radio Occultation was launched from Satish Dhawan Space Center (Sriharikota, India) on board the Indian Remote Sensing OCEANSAT-2 satellite. The Italian Space Agency has established a set of Italian universities and research centers to implement the overall processing Radio Occultation chain, it consists of seven main steps, named Data Generators, execute in series in a specific order. The contribution presents the method for automatic chain processing of Radio Occultation data, from raw data to the characterization of the temperature, pressure and humidity, using a grid computing environment in order to reduce the elaboration time. The grid lets share computing power, databases, and other tools securely across corporate, institutional, and geographic boundaries without sacrificing local autonomy. Our solution allows to make a collaborative framework between research centers, university and distributed data in Italy, where physicists and computer engineers need to work together to process data in efficient way.

**For further details look in the website http://projects.ismb.it/sw-rosa**

The Radio Occultation technique
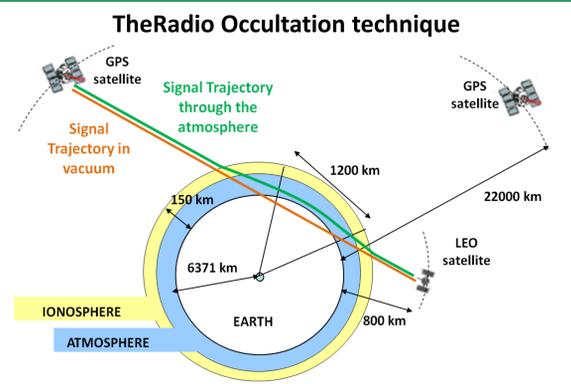
In the framework of this opportunity, the Italian Space Agency funded the development of the overall (and state-of-the-art) RO processing chain which is called ROSA-ROSSA (ROSA-Research and Operational Satellite and Software Activities). The ROSA-ROSSA is integrated in the operational ROSA Ground Segment located in Italy (at the ASI Space Geodesy Center, near Matera) and in India (at the Indian National Remote Sensing Agency, near Hyderabad). The actual version has been implemented on a distributed hardware and software infrastructure exploiting a GRID computing strategy called GPM. The whole processing chain of Radio Occultation (RO) consists of seven main steps, named Data Generators (DGs), these contain several algorithms. All DGs are executed in series with a specific order: the Elaboration step n receives the results of Elaboration step n-1 as input and its output provides the input for Elaboration step n+1, and so on until all the DGs are executed. SWOrD (SoftWare for Oceansat-2 oRbit Determination) acquires RO Data and Navigation Data from Oceansat-2 and support data from IGS and other providers. The input and output files cover a 6h time interval each, for each day we have about 250 files event to process in Grid in sequential way.

ROSA-ROSSA Processing chain

Architecture
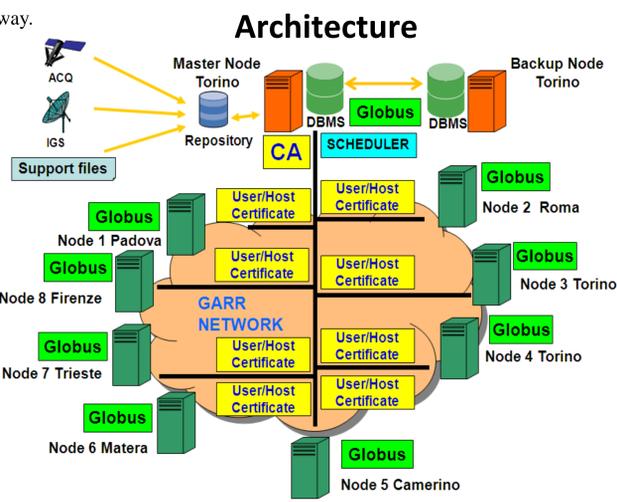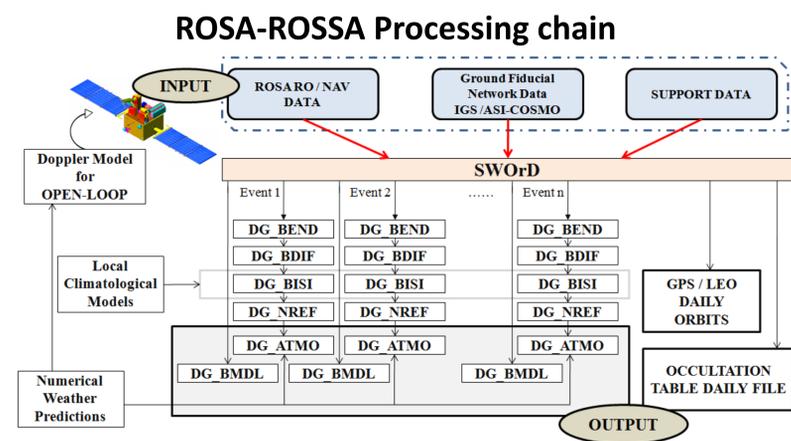
## ROSA-ROSSA Scientific output example: Refractivity Validation
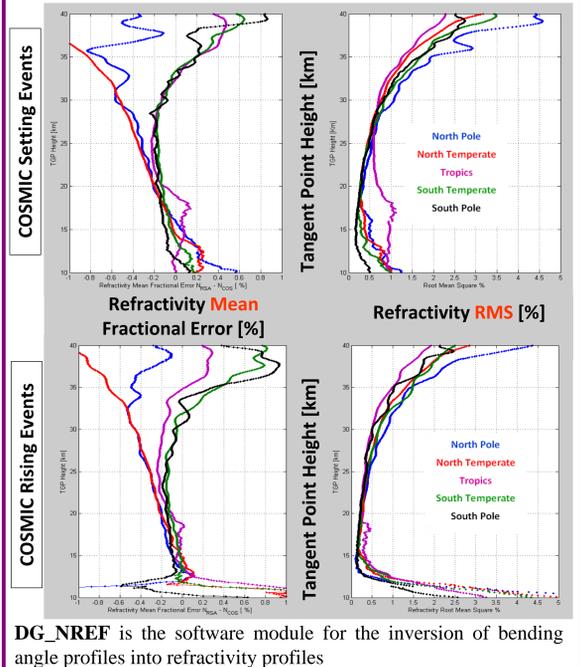
DG_NREF is the software module for the inversion of bending angle profiles into refractivity profiles

Grid Processing Management

The **Grid Processing Management** is an integrated system devoted to handle and process RO data of the OCEANSAT-2 ROSA on board sensor. The general purpose of the project is to share computational resources, transfer a great deal of files and submit jobs from several different organizations of the scientific community located in different places in Italy. It was decided to use the Globus Toolkit middleware, since it allows obtaining a reliable information technology infrastructure that enables the integrated and collaborative use of computers, networks and databases. About the network, we had the opportunity to integrate the nodes into the GARR (the Italian network for research). A Relational Database stores processing chain configuration, information about the location of raw data and products, software needed for execution of each DG, grid configuration and information to write XML files.

The Web application is based on J2EE/AJAX technologies, it allows :
- To monitor the entire processing chain;
- To monitor the grid status (available nodes);
- To download the output files for each Data Generators;
- To upload file for execution in grid environment.
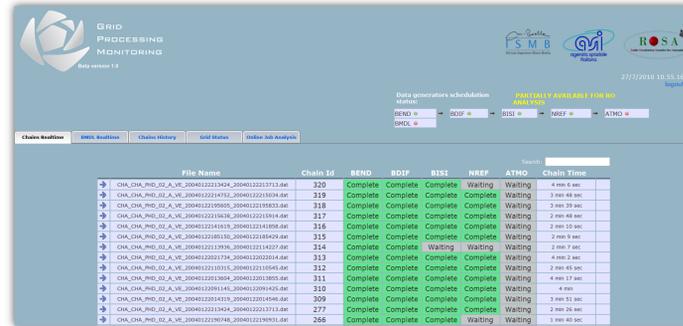
GPM Web Application
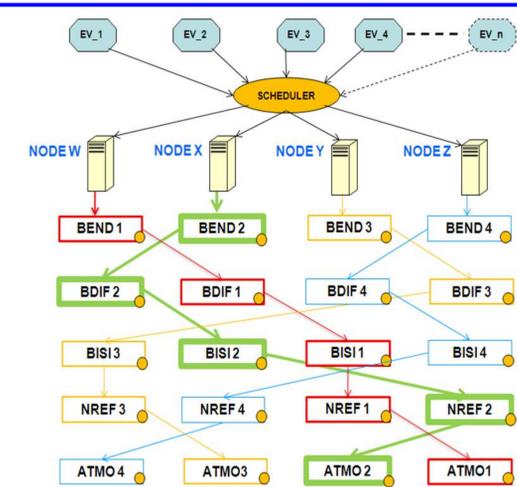
The **Scheduler** functions are split in three phases:

The resources discovery, that receives resources information by the agents installed on each node and selects a list of nodes ready to accept and execute jobs.

The information gathering verifies software and applications needed by the job and software capabilities of each node pre-selected. It depends directly on the job type requested.

The delivery phase checks the pre-selected node list in order to find free nodes and assign to them a new execution.

For this purposes it was developed an ad-hoc scheduler which must take in consideration the following parameters: for the scheduling job polices it checks the nodes capabilities (RAM, CPU and Swap) and for job requirements it checks software, libraries, elaboration time.

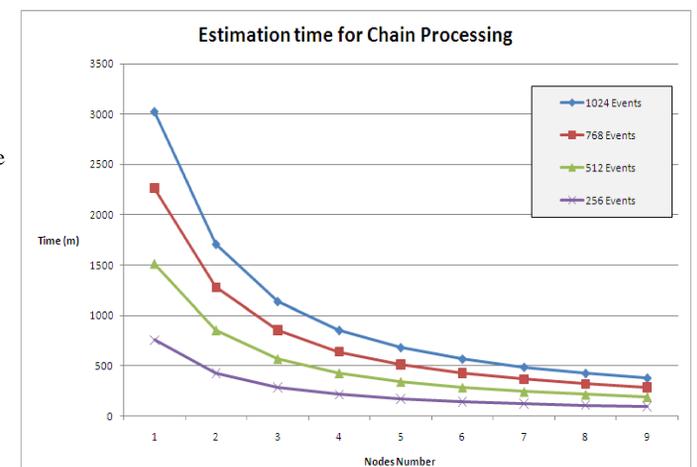## Estimation time for each Data Generator
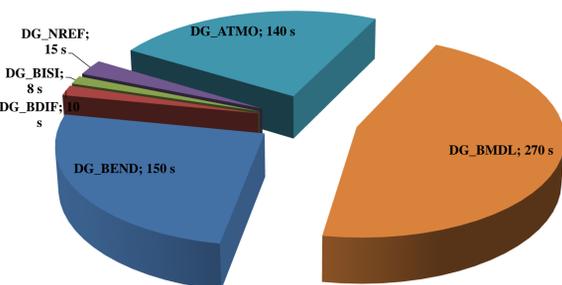
Processing Time

During a learning phase all DGs have been tested at an early stage, considering the number of input, support files needed for elaboration and elaboration time we evaluated the percentage of execution for a single event and for a daily data.

Instead, in a second phase we considered as elaboration time changes by increasing the number of slave nodes and the equation was formulated that represents a estimation time of elaboration and cover both a non-distributed and distributed architecture.

We can notice how time changes when we increase the number of nodes and number of events. We observed that, where N = 9, the total execution time for a day is about 150 minutes, instead of 3021 minutes with N=1.

Distributed Architecture context

$$T_p = \begin{cases} T_s + \dfrac{(T_e * \eta)}{N} & N = 1 \\[2ex] T_s + \beta_s + \dfrac{(T_e + \beta) * \eta}{N} & 1 < N \leq \eta \end{cases}$$

T_p = Total Time Process
T_s = SWOrD Time Process
T_e = Event Time Process
β_s = SWOrD File Transfer Time
β = Total File Transfer Time
η = RO Event Number
N = Grid Node Number

NB: These elaborations have been executed in an single node architecture Dual Core 3.20GHz with 3 GB of RAM. and BLADE architecture Quad Core Xeon 2.50 Ghz with 10 GB of RAM.

NB: VE is able to process raw data acquired during hourly ROSA observation. A 10 min processing time is necessary for the GPM (using 9 nodes) to complete the processing. An extra time of 6 minutes for SWOrD elaboration is needed.

International workshop on Occultations for Probing Atmosphere and Climate (OPAC 2010)