

# Performance of NDVI Index on HPC Resources

Albert G. Saribekyan

Institute for Informatics and Automation Problems of NAS of RA  
e-mail: albert\_saribekyan@ipia.sci.am

## Abstract

Geographic information systems (GIS) [1] are crucial to enable the gathering, analysis, presentation and distribution of spatial and non-spatial data. In some specific GIS applications, such as time-critical simulations or data mining, we need to deal with massive amount of geospatial data storage, retrieval, and processing. The main aim of this article is to analyze and benchmark the performance of Normalized Difference Vegetation Index (NDVI) [2] of satellite images (16 GB [3]) using high-performance computational (HPC) resources, as the developers need to solve the problem of both data and task distribution among serial or parallel environments. The geographical resources analysis support system (GRASS) [4] is used as the main instrument for the study.

**Keywords:** GRASS GIS, High Performance Computing, Remote Sensing, NDVI.

## 1. Introduction

Satellite imagery provides a large amount of useful information and plays a vital role for research developments in Astronomy, Remote Sensing (RS), GIS, Agriculture, Disaster Management and many other fields of study. GIS is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. GIS is used for geographical information science or geospatial information studies to refer to the academic discipline or career of working with GIS. In the simplest terms, GIS is the merging of cartography, statistical analysis, and database technology. These geospatial technologies allow us to monitor natural biological systems. New advances in spatial ecology permit us to put these data in the context of our ecological understanding and to generalize these patterns to advance ecological theories and their applications.

Remotely-Sensed data is one of the most important sources of data for GIS. RS means - acquiring data from a distance usually uses electromagnetic energy sunlight, radar, laser originally captured on photographic film. Recent platforms utilize digital sensors. RS is the observation of an object from a distance. Information about features on the Earth's surface can be gathered from orbiting RS satellites or from a plane (e.g., aerial photography).

Geographic Resources Analysis Support System, commonly referred to as GRASS GIS, is a GIS used for data management, image processing, graphics production, spatial modeling, and

visualization of many types of data. It is a free Software/Open Source released under GNU General Public License  $\geq$  V2. GRASS GIS is an official project of the Open Source Geospatial Foundation. A wide ranging analysis capabilities make it an ideal tool to set up environmental models, as well as to support land planning and management. GRASS was originally developed in the beginning of the 1980s by the US Army Construction Engineering Research Laboratories (USA-CERL), published as a public domain software. The USA-CERL withdrew from the GRASS development in the early 1990s. From 1999 it is developed by an international developer team, which published GRASS as a free software under the terms of the GNU General Public License. GRASS is developed mostly in Linux but binaries for other operative systems are available.

Main GRASS features are:

- 2D raster analysis and 3D voxel (volumes) management
- 2D/3D vector engine with SQL-based DBMS support
- Image processing modules
- Vector network analysis, Linear Referencing System
- Visualization of 2D, 3D maps and volumes
- Interoperable with standard raster and vector formats
- Works on GNU/Linux, Mac OS X, MS-Windows and other POSIX compliant platforms
- Modular architecture and scripting capabilities for batch processing

GRASS GIS is capable of handling raster, topological vector, image processing, and graphic data. It includes more than 350 modules [5] for management, processing, analysis and visualization of georeferenced data. Additionally, processing those data requires a large amount of computation time due to its complex and large processing criteria. This seems a barrier for real time decision making. To switch the job faster, HPC can be a suitable solution. Multi- computer- based distributed systems (clusters and Grids) have a large processing capacity for a lower cost, naturally, choice turns towards developing HPC applications.

Developing GRASS module framework on HPC, three case studies, i.e. “i.vi” to process 13 vegetation indices, “i.lmf” to remove the atmospheric effects from RS images and “r.gaswap” to find out the crop parameters that are not directly visible from RS images, are discussed. The foremost case study (i.vi) discusses the interoperability framework designing issues between the GRASS tool and HPC.

Vegetation Index (VI) is the major set of indicators for vegetation. The GRASS module i.vi, is used to process 13 different vegetation indexes for the satellite images. NDVI (Normalized Difference Vegetation Index) is one of them. The NDVI is calculated from these individual measurements as  $NDVI = (NIR - Red) / (NIR + Red)$ , where RED and NIR stand for the spectral reflectance measurements acquired in the red and near-infrared regions. Fig.1 shows a snapshot of a NDVI output of GRASS Software with RED and NIR bands.

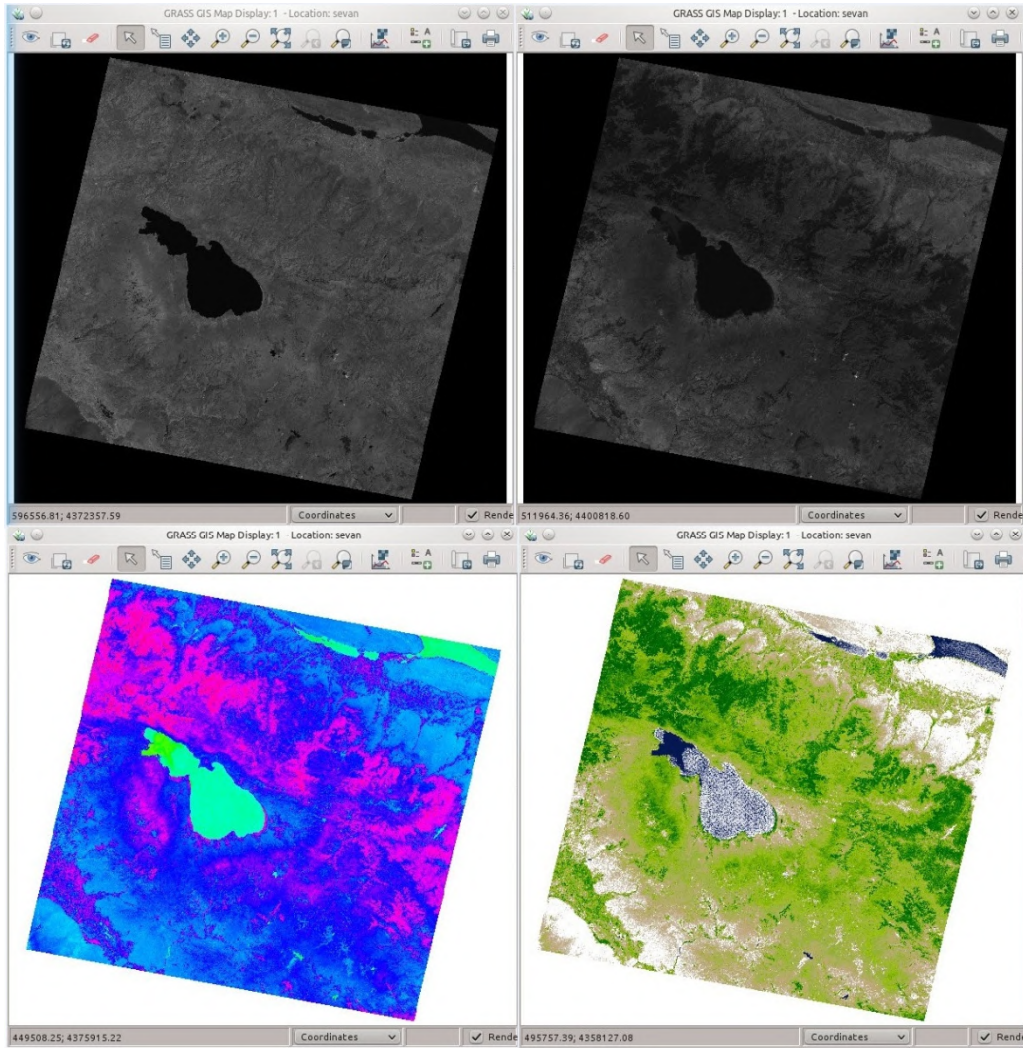


Figure 1: NDVI calculations with GRASS

GRASS module `i.vi` works with raster images (rows x columns). GRASS functions are used to extract row-wise data from the specific band images and store them in buffers. Then, each column value is extracted sequentially from the buffers and sent for generating the specific VI values. Thus, after completing the VI from the row buffers, the row wise VI values are put back into the output image and this process will continue for each row. Fig.2 presents the structure of running `i.vi` module (for simplicity only two band images have been presented).

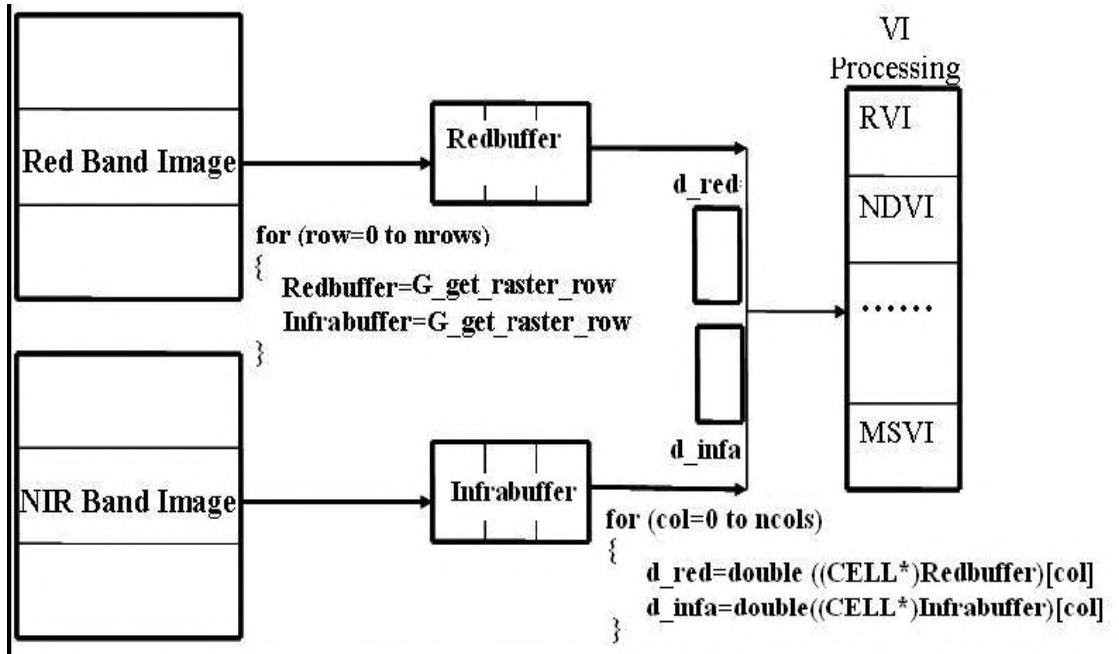


Figure 2: serially running i.vi module structure

In this paper the same task has been compared on two different environments to justify and practically understand the behavior of the parallelized i.vi module of GRASS (which is i.vi.mpi). To achieve this purpose GRASS GIS 7.0.svn [6] was installed on two different platforms. The both HPC have the same hardware but they differ from each other only with the gcc version and operating system. The parallelization was done using MPI standard [7]. Parallelization of i.vi [8] module was illustrated on these platforms using two to eight cores. In fig.3 the computational time has been presented of the same job running on predefined number of cores on two different platforms.

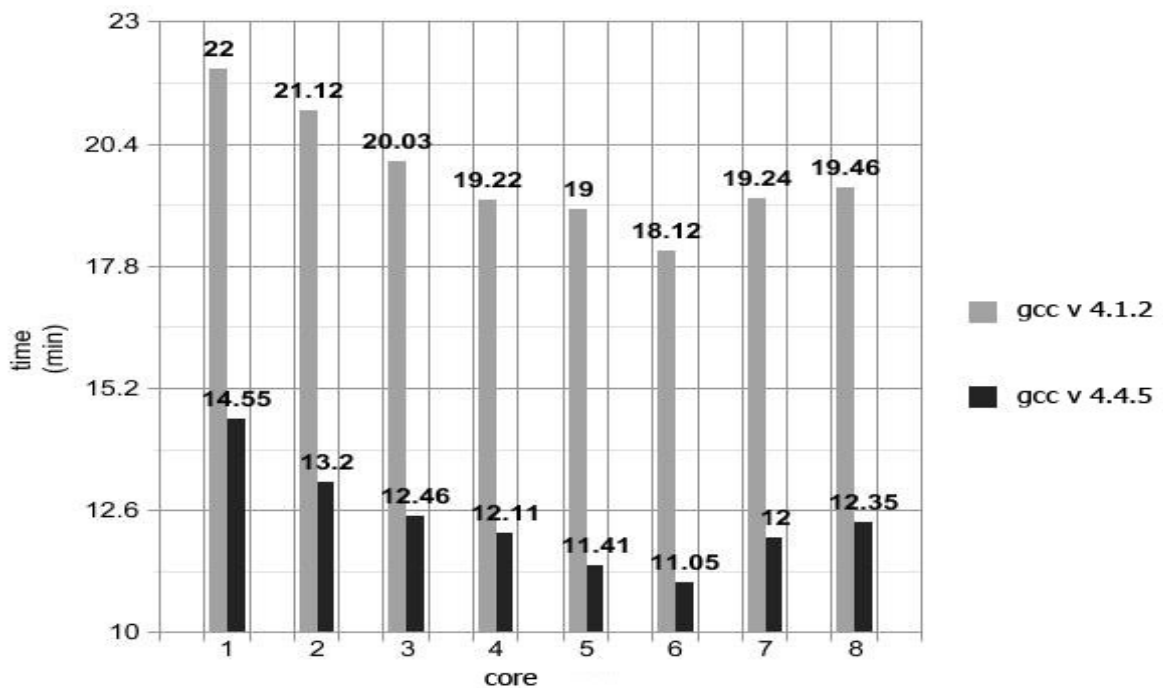


Figure 3: The computational time of job per cores on two platforms

Dark grey scale shows the result of machine in which gcc version 4.4.5 (Debian 4.4.5-8) and operating system is Debian GNU/Linux 6.0 with Linux kernel 2.6.32-5-amd64 and light grey scale shows the result of machine with gcc version 4.1.2 20080704 (Red Hat 4.1.2-50) and operating system is Scientific Linux SL release 5.5 (Boron) with Linux Kernel 2.6.18-274.17.1.el5 x86\_64. These two results obviously show that the i.vi module works best on six cores.

## Conclusion

This research is leading us to be sure that the parallelization of i.vi module is very efficient in terms of the calculation speed, because the same task is done much slower without parallelization. And the second important aspect is the architecture of HPC, especially the version of gcc compiler and the operating system, by which the task solution time has been changed dramatically.

## References

- [1] Geographic information systems as an Integrating Technology: Context, Concepts and Definitions. *Kenneth E. Foote and Margaret Lynch.*
- [2] NDVI index  
<http://gis-lab.info/qa/ndvi.html>
- [3] Database of Landsat satellite images  
<http://glcfapp.glc.f.umd.edu:8080/esdi/>
- [4] Neteler, M. and Mitasova, H., Open Source GiS: A GRASS GIS Approach. Second Edition, 2003.  
*Kluwer Academic Publishers.*
- [5] GRASS GIS 7.0.svn Reference Manual  
[http://grass.osgeo.org/grass70/manuals/full\\_index.html](http://grass.osgeo.org/grass70/manuals/full_index.html)
- [6] GRASS GIS official website  
<http://grass.osgeo.org/>
- [7] The Message Passing Interface (MPI) standard:  
<http://www-unix.mcs.anl.gov/mpi/>
- [8] Parallel GRASS jobs  
[http://grasswiki.osgeo.org/wiki/Parallel\\_GRASS\\_jobs](http://grasswiki.osgeo.org/wiki/Parallel_GRASS_jobs)

**Submitted 10.11.2012, accepted 04.02.2013.**

Բուսականության նորմավորված տարբերության ինդեքսի  
հաշվարկման արտադրողականությունը բարձր  
արտադրողականության հաշվողական ռեսուրսներում

Ա.Սարիբեկյան

Ամփոփում

Երկրատեղեկատվական համակարգերը (ԵՏՀ) կարևորվում են տարածական և ոչ տարածական տվյալները խմբավորելու, վերլուծելու, ներկայացնելու և

տարանջատելու համար: Որոշ տեսակի ԵՏՀ ծրագրային համակարգերում, ինչպիսիք են կարճ ժամկետներում մոդելավորման կամ տվյալների ձեռքբերման համակարգերը, գործ ենք ունենում մեծ քանակի տարածական տվյալների պահպանման, որոնման և մշակման հետ: Այս հոդվածի հիմնական նպատակն է հետազոտել և վերլուծել տիեզերական պատկերների բուսականության նորմավորված տարբերության ինդեքսի հաշվարկումը՝ օգտագործելով բարձր արտադրողականության հաշվողական համակարգեր, քանի որ հետազոտողների անհրաժեշտ է լուծել առաջացող խնդիրները միաժամանակ զուգահեռ և ոչ զուգահեռ միջավայրերում: Ուսումնասիրությունների համար որպես հիմնական գործիք ընտրված է աշխարհագրական ռեսուրսների վերլուծության ապահովման համակարգեր ծրագրային փաթեթը:

**Производительность вычисления нормализованного относительного  
индекса растительности на высокопроизводительных  
вычислительных ресурсах**

А.Сарибекян

**Аннотация**

Геоинформационные Системы имеют решающее значение для сбора, анализа, представления и распределения пространственных и атрибутивных данных. В некоторых ГИС приложениях, таких как срочное моделирование или анализ данных, мы имеем дело с хранением, поиском и обработкой огромного количество геопространственных данных. Основной целью данной статьи является анализ и тестирование производительности вычисления нормализованного относительного индекса растительности из спутниковых изображений, используя высокопроизводительные вычислительные ресурсы, так как разработчики должны решать задачи в параллельных и не параллельных средах. В качестве основного инструмента для исследований использован программный пакет системной поддержки анализа географических ресурсов.