

Satellite image time series analysis with the sits package

Gilberto Camara¹, Victor Maus², Luiz Fernando Assis¹, Rolf Simoes¹, Adeline Maciel¹,
and Gilberto Ribeiro de Queiroz¹

¹*Image Processing Division, National Institute for Space Research (INPE), Av dos Astronautas
1758, Sao Jose dos Campos, 12227-001 Brazil.*

Email: {gilberto.camara}@inpe.br

²*International Institute for Applied System Analysis, Schlossplatz 2, 2631 Laxenburg, Austria*

A satellite image time series is...

first, get information about the WTSS (web time series service) see WTSS paper for more information ("Web Services for Big Data") which shows the time series for the "evi" band

1 Introduction

Earth observation (EO) satellites produce vast amounts of geospatial data. The Landsat archive holds over five million images of the Earth's surface, with about 1 PB of data. New satellites from Europe, USA, China, Brazil, and India generate yearly as much data as one Landsat satellite in a decade. Most space agencies have adopted an open data policy, making unprecedented amounts of satellite data available for research and operational use. This data deluge has brought about a major challenge for Geoinformatics research: *How to design and build technologies that allow the EO community to analyse big data sets?*

When scientists use big EO data they face the burden of organising thousands of files, downloaded from the space agencies archives. To manage such large data sets, researchers need stable and efficient solutions that support their analytical tasks. To choose a solution is hard because the technology for large data handling and analytics is evolving. Alternatives include MapReduce-based solutions such as Google Earth Engine [Gorelick2012], object-relational DBMS extensions such as Rasdaman [Baumann1998] and distributed multidimensional array databases such as SciDB [Stonebraker2013]. Since each of these architectures takes on a different approach, understanding the benefits and drawbacks of each one helps researchers choose what best fits their needs.

```

> library(sits)
> URL <- "http://www.dpi.inpe.br/tws/wtss"
> inpe <- sits_configWTSS (URL,
+                           coverage = "mod13q1_512",
+                           bands = c("ndvi", "evi", "nir"))
> long <- -55.51810
> lat <- -11.63884
> series.tb <- sits_getdata(longitude = long, latitude = lat, wtss = inpe)
> series.tb %>%
+   sits_select (bands = "evi") %>%
+   sits_plot ()

```

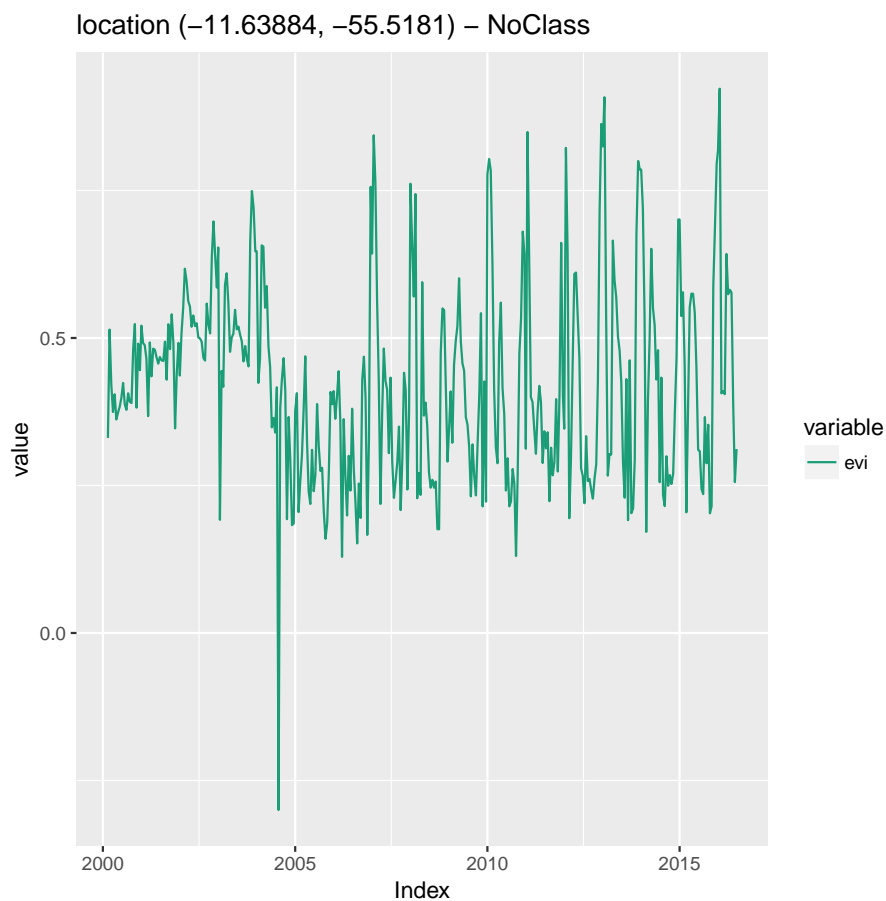


Figure 1: Time series for EVI band

Given the diversity of options, researchers would gain from documented experience that helps them to assess how proposed big data architectures fit the needs of geospatial data analysis. Recent papers describe algorithms required for EO analysis [Vatsavai2012] [Nativi2015] and report case studies using specific architectures [Planthaber2012][Kroger2012]. However, to make progress on big geospatial data analysis, we need to engage the large community of remote sensing researchers. In this paper, we consider how big EO data architectures can support the needs of data analytics. Our paper examines ways to cut the effort required for researchers to develop and validate algorithms for extracting information for big EO data.

We take the viewpoint that architectures should serve applications, and not the other way around. To clarify the researcher's problem, we consider the needs for an important EO application: land use and land cover change (LUCC) analysis. We propose an architecture based on open-source tools that combines array databases with statistical analysis. We evaluate this design to assess how it meets the needs of EO scientists and compared with other approaches that aim to meet these needs. This paper also puts forward a set of criteria to build researcher-friendly architectures for big EO data analysis.

references-esensing.bib