

DEVELOPMENT OF A WEB SERVICE AND ANDROID 'APP' FOR THE DISTRIBUTION OF RAINFALL DATA. A BOTTOM-UP REMOTE SENSING DATA MINING AND REDISTRIBUTION PROJECT IN THE AGE OF THE 'WEB 2.0'.

Vasco M. Mantas⁽¹⁾, A. J. S. C. Pereira⁽¹⁾, Zhong Liu⁽²⁾

(1) IMAR, Department of Earth Sciences, University of Coimbra, Portugal, Email: vasco.mantas@dct.uc.pt

(2) CSISS, George Mason University and NASA GES DISC

ABSTRACT

A project was devised to develop a set of freely available applications and web services that can (1) simplify access from Mobile Devices to TOVAS data and (2) support the development of new datasets through data repackaging and mash-up.

The bottom-up approach enables the multiplication of new services, often of limited direct interest to the organizations that produces the original, global datasets, but significant to small, local users. Through this multiplication of services, the development cost is transferred to the intermediate or end users and the entire process is made more efficient, even allowing new players to use the data in innovative ways.

Keywords: TRMM, rainfall, Data distribution, python

1. INTRODUCTION

The rapid evolution of the internet and the dawn of the so-called web 2.0 [1] led to a rapid increase of readily available data, including numerous remote sensing products. This is also due to the increasing number of Earth Observation (EO) missions and the opening of several data archives. For these reasons, the volume of data became such that automated data mining systems are now a necessity [2]. In fact, the opening of the data archives has been credited with a sensible increase in the number of scientific papers and applications produced on the basis of such data, of which the Landsat archive of the USGS is a good example [3].

However, for the end-user, it is often more convenient to access the information not in the typical raster format but in other, more easily manageable outputs, such as text files. This is even more pressing when the objective is to create

expedite visualizations or when the available bandwidth and computing capabilities limits the use of large, multiband raster files.

A successful example of such a data distribution system is offered by the TRMM Online Visualization and Analysis System (TOVAS) [4], one of the instances included in the GIOVANNI tool of NASA's Goddard Space Flight Center [5].

The increasing demand for Rainfall data, such as that provided by TOVAS is yet another argument in favor of these data distribution systems. The datasets and TOVAS itself have been used in a wide range of fields [6] including public health (e.g. [7]). The potential users of these data are often not remote sensing experts nor programmers and the simplified formats in which the information is provided is an enormous advantage to facilitate the penetration of such technologies in new fields. There are also undeniable advantages in the use of TOVAS-like data in educational contexts, especially in k-12 and undergrad curricula [8]. Users can focus on analyzing the data, bypassing the hurdles associated to the processing and handling of large volumes of information.

Nonetheless, and despite the aforementioned advantages, individual instances at GIOVANNI offer information on a single family of products (rainfall estimates, ocean color, aerosols, etc...) and it is up to the user to download the different datasets and compile them offline. But taking into account that there are several other data distribution portals (even when considering NASA alone), with information that can supplement that of GIOVANNI, the purpose of facilitating the immediate visualization is made more difficult. The full reach of near real time applications can only be appreciated through the development of automated data analysis systems namely in the case of Early Warning Systems (e.g. [9]).

In this context, a project was devised to create a set of freely available applications and services, under

a common framework that can: 1) simplify access to TOVAS data from mobile devices and 2) support the development of customized datasets through data repackaging and mash-up. Collectively, the set of applications and services is known as the TRMM EXPedite ExpLoration Of Rainfall Estimates ('TRMM Explorer').

2. THE MOBILE APP

To simplify the development process and contribute to a more efficient workflow, the solutions of the TRMM Explorer (TE) project were developed by order of complexity.

As such, the first data access solution created for the TE was an application for mobile devices running on Android (Fig. 1).

The main objective of the application (or 'app') was to emulate the standard TOVAS data access portal, but in a simplified manner. Such an application ought to address the constraints inherent to mobile devices including the small screen size and resolution, limitations to complex text inputs, processor power, limited memory and battery life [10]. Additional requirements included the need for the user to be able to take advantage of the built-in sensors of the device, especially the GPS (to set the extents of the data request). By taking advantage of the ASCII output capabilities made available by TOVAS, a first app was designed and deployed successfully. With a simple User Interface (UI) and named TRMM.Mobile, it offers users a simple way of accessing Rainfall Estimates on the fly from mobile devices (Fig. 1). One of the objectives of the TRMM.Mobile application is to offer both the development team and the user community a reliable base from which other, more advanced applications can be developed.

The app was designed in a 'touch-and-go' perspective, in which the user is not required to manually insert text.

The Area of Interest (AOI) can be set once the GPS acquires the local coordinates. The set of coordinates, or those inserted manually, can also be stored in a simple local database for later use.

The start and end date of the period to be analyzed can also be selected using a simple selection system with buttons controlling the process of inserting the day, month and year. In seconds, a user can set all the required parameters and request the data, which is then displayed on the screen.

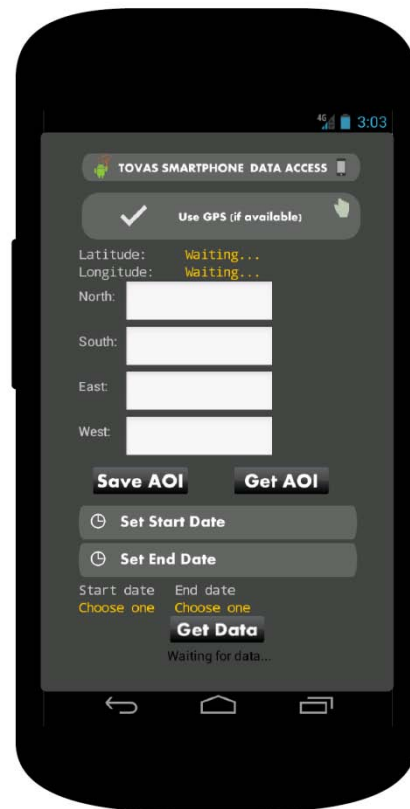


Figure 1. Screenshot of the TRMM.Mobile at work in an android emulator.

3. THE WEB SERVICE

The TRMM Mobile application, despite the straightforwardness of its operation still lacks important capabilities useful to the user community. In particular it renders impossible, or impractical, complex calculations and precludes desktop visualizations. It was thus considered the development of a web-based service capable of retrieving TOVAS data but with the capability of creating new, derived products.

Analyzing data for pre-defined regions, resampling the data to new temporal resolutions or calibrating it against *in situ* measurements are some of the declared goals. To accomplish such objectives it was necessary to develop a program capable of handling user requests and, at the same time, accommodate pre-defined services. The format of the output files and visualizations should also be flexible, enabling the users to access it as the final product or use it as the input to other applications. At the same time, and although TOVAS is at the center of the project, it was deemed pertinent to integrate other products in the system, including

MODIS data or discharge measurements from USGS river gauges.

Meeting all these requirements was challenging and led to the selection of python as the language in which the system would be developed.

Python as an object-oriented, extensible and high-level programming language is especially suitable to projects such as the TRMM Explorer. There are also numerous freely available modules that simplify the overall development effort, making of it a popular choice in scientific projects such as this [11].

The project includes a series of programs based on a common general data fetching and analysis application. Each specific program is meant to address certain requirements (e.g. data fetching from different servers) and can be easily integrated into a single application offering the flexibility that is sought after with this project.

In this paper we present several examples of applications developed in the context of the TRMM Explorer, some of which already available online.

3.1. Southern Angola Rainfall Monitoring

The development of research projects in Southern Angola by the development team created the setting for a live test of the system while enabling its continued development and improvement.

The scarcity of *in situ* rain gauges in Southern Angola together with the environmental challenges caused by changing land cover, a growing demand for water and recurrent flooding and drought events motivated the development of several monitoring instances. These TRMM Explorer instances are focused on specific river basins and retrieve data from several TOVAS products which are then repackaged and displayed in a public website or sent via email or to mobile apps.

Although the instances are functional for several rivers, public release shall happen gradually, in response to regular evaluations of system performance and user feedback.

There are several challenges to the use of data archived in systems such as TOVAS in developing nations. The language barrier, limited internet connectivity and interactivity are important issues

that prevent potential users from accessing the information and use it regularly. Amongst the options included in the TRMM Explorer program is language localization, made possible through a simple configuration file. However, localization efforts in this context may require an effort beyond the simple translation of contents into one or more local languages [12]. It is indeed necessary to adapt the entire system to local practices and expectations. At the same time, the overall system must remain flexible enough to prevent the development of countless instances, similar in objective but powered by several variations of the same code, which would lead to a maintenance and management nightmare.

To meet these goals, the rainfall estimates are delivered to the users by means of interactive charts powered by the Google Visualization API which can be embedded in a website. Other possible outputs include email, tables and maps (in kml).

The data is fetched using a URL constructed from the parameters partly stored in a pre-defined configuration file, the 'Instance Configuration File' (ICF) (Fig. 2). The ICF, which is a python file, contains a set of variables including the coordinates of the area of interest and the weight to be assigned to a specific TOVAS cell (used in the weighted sum of the daily rainfall estimates for a specific river basin). When no weight is required the value can simply be assigned to '1' in all cells. In this example, the weights were calculated using ArcGIS 10.1. The TOVAS 0.25° grid was overlaid to SRTM-derived basins, delineated using standard topographic analysis. The area of each cell containing a portion of the basin was then calculated, providing the weight of the given grid element.

The daily rainfall data thus retrieved from TOVAS by the TRMM Explorer Python program is then stored in an html file which generates the visualization. This file is cached and displayed whenever the specific instance is called, with advantages to the users (faster access to the data) and to the system which no longer needs to handle individual requests.

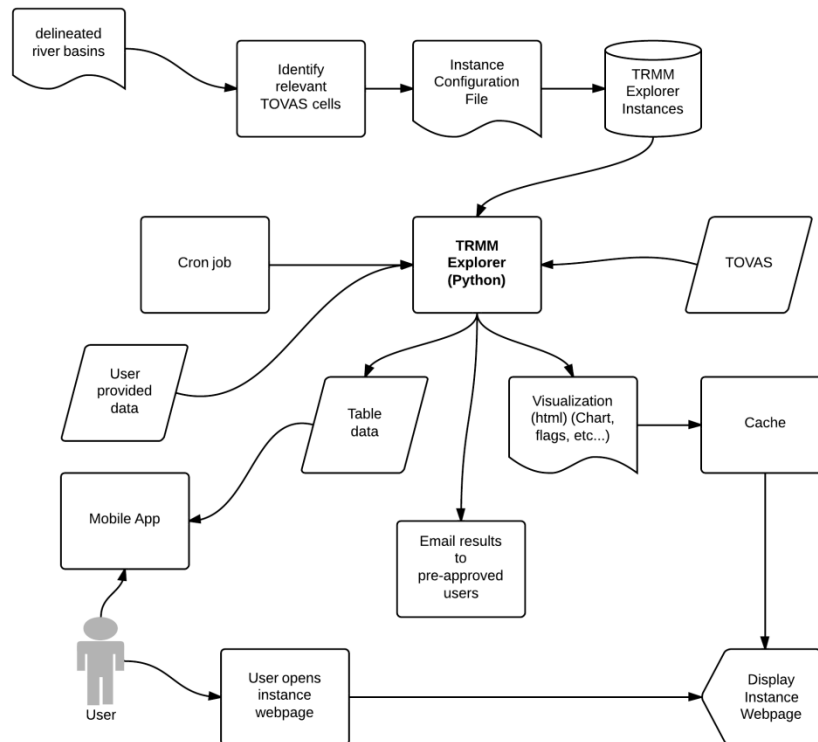


Figure 2. The typical workflow of the automated Southern Angola rainfall monitoring instances.

Although this approach may not be applicable in other cases, including when the user must have full control over the location and timeframe for which the data must be fetched it is reasonable for this and other, similar applications.

The web-based component of the TRMM Explorer is focused on the deployment of scalable, continuous monitoring solutions for specific areas making the aforementioned approach reasonable and cost-effective in this context. The automated analysis and visualization generation is scheduled to take place on a regular basis with Cron jobs assigned according to the product being analyzed (3-hours, daily, monthly). The system can also be configured to send automated email messages to predefined recipients containing a summary of the estimates.

3.2. 'Active' TE-driven mobile applications

The development of the online python program paved the way to create more complex 'apps' meant to be run on mobile devices.

Such apps can display the output of the python programs in a way that is compatible with the already mentioned limitations of smartphones. Built on top of the application described in section 2, a multi-product application using a simple user interface was created to retrieve local information

from multiple sources. The app creates a URL that is sent to a TRMM Explorer instance. The URL contains a query string including the coordinates of the site retrieved from GPS (e.g. <http://TEhost/instancename/?lat=value&lon=value>). The python program then fetches data from TOVAS (daily, Real Time) and Normalized Difference Vegetation Index (NDVI) data from subsets of the MODIS Land Products made available by the ORNL (ORNL DAAC [13]). The MODIS data is actually available through an online python client, making it easy to fetch the desired information. All the data are then served through an html page containing both products. This page, containing easy to read tables, can be displayed seamlessly in the UI of the app.

Although we only included rainfall estimates and NDVI data in this example, other products can be integrated, effectively creating the conditions for a 'field companion', ideal in the support of field surveys and educational activities.

3.3. Customizable requests

The TRMM Explorer was also designed to accommodate customized data requests. The requests can be controlled by the ICF or by string queries. The capacity to do so was implemented

upon the development of the Rainfall/NDVI mobile application which required string queries in the URL as aforementioned. A number of queries are currently possible including the coordinates of the area of interest as well as the start and end date. The user can also control the values used in an equation of the type “ $mx + b$ ”, with x being the original rainfall estimate and therefore obtaining derived values. For this, it is necessary to include the slope (m) and intercept (b) in the query.

We present an example (fig. 3) of such a case in which *in situ* rainfall gauge data was compared against TRMM monthly rainfall estimates (3B43 V7). A simple linear regression enables an approximate calibration of the satellite estimates for this area. Inserting the resulting equation in the query enables users to visualize and retrieve the converted data immediately.

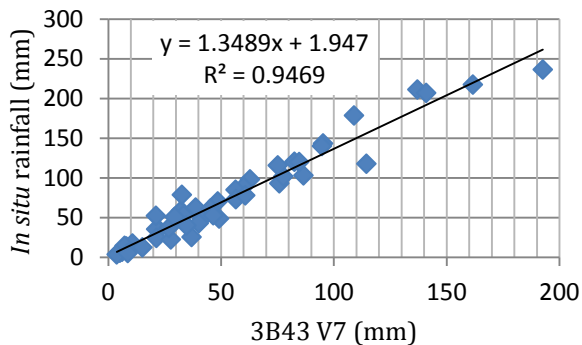


Figure 3. Correlation between *in situ* rainfall measurements in Reveles, Portugal (SNIRH) and Monthly 3B43 V7 data extracted from TOVAS. The custom request to visualize the converted data would be:

<http://www.trmmexplorer.org/instancename/?lat=40.12&lon=-8.74&m=1.3489&b=1.947>

Other parameters can also be defined providing there is interest in the community for such capabilities.

3.3. Historical datasets

Historical rainfall events can be used in education as case studies. Tools making such datasets available can offer the academic community

important opportunities for hands-on learning. By taking advantage of the capabilities of the system, the extraction and storage of select educational datasets was implemented, in order to illustrate exceptional rainfall (or drought) events. These datasets can be supplemented by information from sources other than TOVAS. Once retrieved and validated, the datasets are stored to minimize resource consumption. The data can thus be obtained easily at the project website in an ever growing archive of visualizations. Educators may also request specific teaching materials, consisting of rainfall estimates and other data, displayed in the standard TRMM Explorer formats but for custom regions and dates. The datasets may even be used in the development of TE-compatible applications given that the outputs are consistent with those returned by the standard instances.

For illustration purposes, an example from an historical rainfall event in Tennessee, USA is presented.

In May 2010, record-breaking, concentrated rainfall resulted in extensive flooding across Tennessee, Kentucky and Mississippi, leading to the loss of life and property.

A customized visualization (fig. 4) was generated to depict this event. Data was fetched from TOVAS and the USGS river gauge located at the Obion River, near Greenfield (ID 07024500) using the TRMM Explorer system. Additionally, and using the Visualization Configuration File (VCF, not to be confounded with the ICF), additional datasets were added, including reports, flags and links to Landsat imagery. The VCF is a customizable text file that can either be stored at the TE server or at a third-party host and contains ancillary information to be added in specific dates. The TE program then identifies the dates for which there is additional information and adds it as annotations to the visualization. Information may include imagery, data quality flags or *in situ* measurements to name a few examples.

Using the TE, it is possible to download the visualization and use it even if in an offline context although links to imagery or other web-based data will be unavailable unless there is a working internet connection.

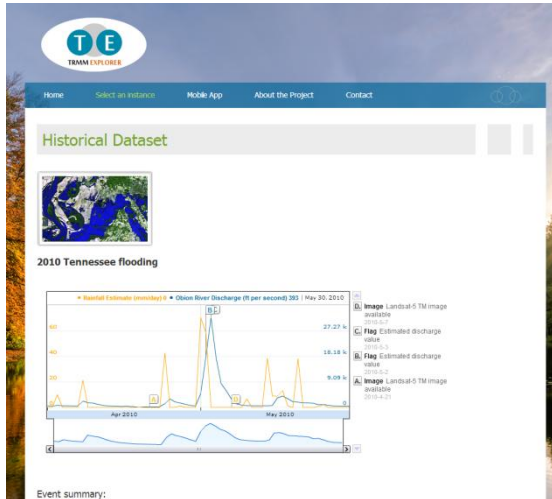


Figure 4. The 2010 Tennessee floods in a customized, dynamic visualization at the website of the TRMM Explorer.

The historical datasets are, in summary, a static implementation of the standard TRMM Explorer capabilities.

4. DISCUSSION AND FUTURE PLANS

The TRMM Explorer is still under development and it is expected to evolve rapidly once it becomes available to the public. We believe that the functionality presented in this paper provides the blueprint for future versions to be released in the near future.

In summary, the TRMM Explorer accommodates three types of datasets: dynamic, customizable and archived. The archived datasets are the easiest to maintain. These are static, cached files containing historical datasets relevant for research or educational purposes. Dynamic datasets are regularly updated (i.e. every day for daily products) and are configured through an ICF. Customizable datasets are most complex and resource-intensive. The servers must be able to accommodate an irregular number of requests, possibly simultaneous and with varying complexities. Given that the objective of the TRMM Explorer is not to replace TOVAS but to simplify access to specific types of data and generate new forms of displaying it, the customizable requests are also the least relevant. A notable exception to this principle would be the development of embeddable visualizations of TOVAS data, using calibration factors derived from *in situ* data. In fact, the aforementioned mobile application (section 3.2) represents an

example of a customizable request, but one that is handled by the app and not the user.

There are still several challenges before both the mobile application and the online instances can become fully available and considered stable. The mobile application still lacks a simple method to enable the user to export the TOVAS data for later use. It is also designed to retrieve data from daily data (3B42 V7 derived) but other products could be interesting as well.

At the same time it is important to emphasize that the TRMM Explorer-powered 'app', fetching rainfall estimates and NDVI is displaying the TE-recalculated values in order to match these with the time intervals (16-day) of the standard MODIS data. Despite this improvement, when compared with the original, 'passive' app described in section 3.1, more efficient data display solutions are still warranted to maximize field usability and minimize power consumption.

To achieve such improvements, the feedback from the users is pivotal. The webportal that was created to support the project and distribute the information and solutions will become the focal point for user-TE team interactions. The role of the social media must also be equated, especially in the case of education-focused tools.

It was also created the figure of the 'trusted user' that is granted special privileges when it comes to configure customized instances and add ancillary data to the visualizations through the VCF file. The selection process is based on the evaluation of the requests submitted by the users for the establishment of new instances. This is a straightforward process, described at the project's website and requiring the submission of a 1-page project describing the relevance of the instance to the public or the scientific community (in English, Portuguese or Spanish).

Future developments may also address the significant interest in the development of scripts that may connect the TRMM Explorer to desktop GIS or entirely replicate its functionality. In particular, scripts for ArcGIS and Quantum GIS are in the early development stages.

But if future developments will be focused on addressing user demands such an expansion of the TE project must be sustainable in order to guarantee that the foreseeable growth of instances can be accommodated without a deterioration of the service. One possible path towards sustainability could include the release of the source code or

refocus the program so that it app development is transferred to other groups or even to the users. The abundance of simple mobile app programming systems can transform almost anyone into a developer (e.g. [14]). As such, it may be reasonable to develop efficient systems capable of serving data to a nowuser-developed app, instead of both systems simultaneously.

By taking advantage of publically available data, it was possible to create a new data distribution system built around the concept of creating an enhanced data visualization experience for the user both on the desktop and mobile devices. The market pervasiveness and growing power of smartphones enabled this concept to be implemented and move from simple emulators of web portals to real data mining tools and field aids. The web component of the project opens way to the establishment of low-cost rainfall monitoring systems for a wide range of applications on both a global and local scale.

ACKNOWLEDGEMENTS

This work was financially supported by the Science and Technology Foundation (FCT, grant SFRH/BD/89972/2012).

6. REFERENCES

1. O'Reilly, Tom (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications & Strategies*, 65, pp. 17 - 37.
2. Yue, P., Di, L., Wei, Y., Han, W. (2013). Intelligent services for discovery of complex geospatial features from remote sensing imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, 83, pp. 151 – 164.
3. Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat, *Remote Sensing of Environment*, 122, pp. 2 – 10.
4. Liu, Z., Ostrenga, D., Teng, W., and Kempler, S. (2012). Tropical Rainfall Measuring Mission (TRMM) Precipitation Data and Services for Research and Applications, *Bulletin of the American Meteorological Society*, doi: <http://dx.doi.org/10.1175/BAMS-D-11-00152.1>
5. Berrick, S.W., Leptoukh, G., Farley, J.D., and Rui, H. (2009). Giovanni: A Web service workflow-based data visualization and analysis system, *IEEE Trans. Geosci. Remote Sens.*, 47, 1, pp. 106-113.
6. Liu, Z., Rui, H., Teng, W. L., Chiu, L. S., Leptoukh, G. G., & Vicente, G. A. (2007). Online visualization and analysis: A new avenue to use satellite data for weather, climate and interdisciplinary research and applications. *Measuring Precipitation from Space - EURAINSAT and the future, Advances in Global Change Research*, 28, 549-558.
7. Liu, Z., L. Chiu, W. Teng, H. Rui, and G. Serafino, (2002). TRMM rainfall products and tools for tropical infectious disease studies, 15th Conference on Biometeorology/Aerobiology and 16th Congress of Biometeorology, Kansas City, MO.
8. Passow, M. (2010). TRMM - Tropical Rainfall Measuring Mission: Bringing remote sensing of precipitation into your classroom, *Terrae didat.*, 6 (1), pp. 03-08.
9. Moffitt, C., Hossain, F., Adler, R., Yilmaz, K., Pierce, H. (2011). Validation of TRMM-based global Flood Detection System in Bangladesh, *International Journal of Applied Earth Observation and Geoinformation*, 13, pp. 165 – 177.
10. Weng, Y., Sun, F., Grigsby, J. (2012) GeoTools: An android phone application in geology, *Computers & Geosciences*, 44, pp. 24 – 30.
11. Oliphant, T.E. (2007). Python for scientific computing, *Comput. Sci. Eng.*, 9, pp. 10-20.
12. Westland, J. (2002). The cost of errors in software development: evidence from industry, *The Journal of Systems and Software*, 62, pp. 1-9.
13. Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC)

(2011). MODIS subsetting land products, Collection 5. Available on-line [<http://daac.ornl.gov/MODIS/modis.html>] from ORNL DAAC, Oak Ridge, Tennessee, U.S.A. Accessed August, 2013.

14. Turbak, F. (2012). Blocks languages for creating tangible artifacts, *Visual Languages and Human-Centric Computing (VL/HCC)*, 2012 IEEE Symposium, Sept. 30 2012-Oct. 4 2012.