

# The Earth Observation Time Series Analysis Toolbox (EOTSA) - An R package with WPS, Web-Client and Spark integration

Ulrich Leopold<sup>1</sup>, Benedikt Gräler<sup>2</sup>, Henning Brede<sup>2</sup>, J. Arturo Torres-Matallana<sup>1</sup>, Philippe Pinheiro<sup>1</sup>, Mickaël Stéfas<sup>1</sup>, Thomas Udelhoven<sup>3</sup>, Jeroen Dries<sup>4</sup>, Bernard Valentin<sup>5</sup>, Leslie Gale<sup>5</sup>, Philippe Mougnaud<sup>6</sup>, and Martin Schlerf<sup>1</sup>

<sup>1</sup> Luxembourg Institute of Science and Technology (LIST), Environmental Research and Innovation, Esch-sur-Alzette, Luxembourg (ulrich.leopold@list.lu)

<sup>2</sup> 52°North Initiative for Geospatial Open Source Software GmbH, Germany

<sup>3</sup> Environmental Remote Sensing and Geoinformatics, Trier University, Germany

<sup>4</sup> Flemish Institute for Technology & Research (VITO), Centre for Remote Sensing and Earth Observation Processes, Belgium

<sup>5</sup> Space Applications Services NV/SA, Belgium

<sup>6</sup> European Space Agency, ESRIN, Italy

The motivation behind the development of the Earth Observation Time Series Analysis (EOTSA) Toolbox is to ease the access and analysis of data provided by the PROBA-V mission exploitation platform (MEP) at VITO. While designing the tool box, we envisioned a number of different application scenarios. The R package itself can be used locally or within the VITO infrastructure on dedicated virtual machines. This enables the user to directly customise or extend the tools in R as needed. Furthermore, the WPS4R wraps the functionality of EOTSA as a Web Processing Service (OGC WPS). The user can use dedicated workflow modelling tools (in this project: the Automatic Service Builder developed by Space Applications) or some Geographic Information Systems (GIS) to access and integrate the WPS. With the help of a map-based Web Client, the user directly executes the WPS processes that analyse the raster data time series for a given bounding box on VITO's Spark cluster. Results are provided as a Web Coverage Service (OGC WCS) and a Web Mapping Service (OGC WMS) based on GeoServer for further use in any GIS.

The Earth Observation Time Series Analysis Toolbox is an R development within the PROBA-V MEP Third Party Services project and has been largely funded by the European Space Agency. We developed the toolbox according to the needs of the time series analysis community. It is optimized for the data provided by the PROBA-V mission exploitation platform (MEP) at VITO. The analytics tools largely build on existing specialized R packages and functions for raster and time series analysis. In the following paragraphs, we will present different examples of how the toolbox can be used. Readers interested in the technical aspects will find details on the system design at the end of this display.

## Example: Summary Statistics

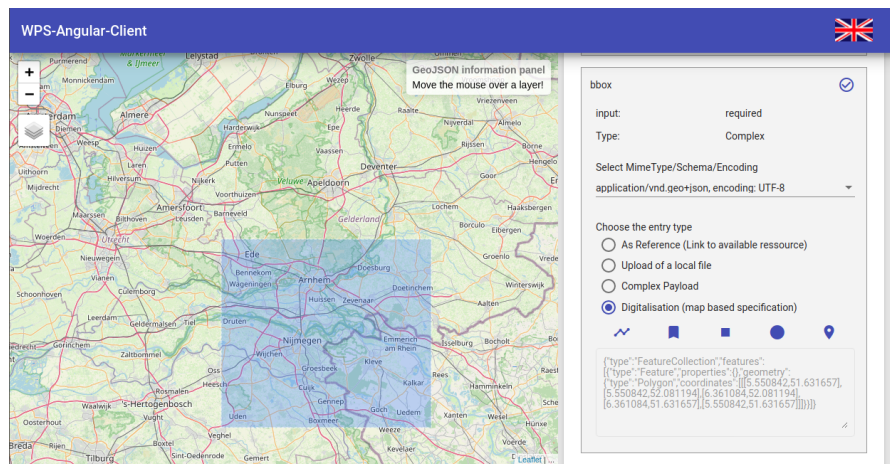


Figure 1: Screenshot of the Web client for process selection and parametrisation.

A starting point for many studies is getting to know the data. As such, the EOTSA toolbox provides rasterized functions that calculate 5 point summary statistics per pixel time series. Figure 1 illustrates the process selection and parametrisation in the Web client. In addition to defining the spatial bounding box, the user must also specify the temporal window.

The user can unfold an execution tab and submit the request with its parameters to the WPS on the back-end. On the server side, a R session is initiated in the Spark cluster. This directly loads, clips and distributes the spatio-temporal raster data according to the process parameters. After all worker nodes in the cluster have applied the dedicated analysis (5 point summary statistics in this example) and the results have been re-combined, a GeoServer instance imports and publishes the results. Once the data has been processed and made available, the user can directly view it in the Web client (Figure 2), or import it into a GIS via WMS and WCS interfaces.

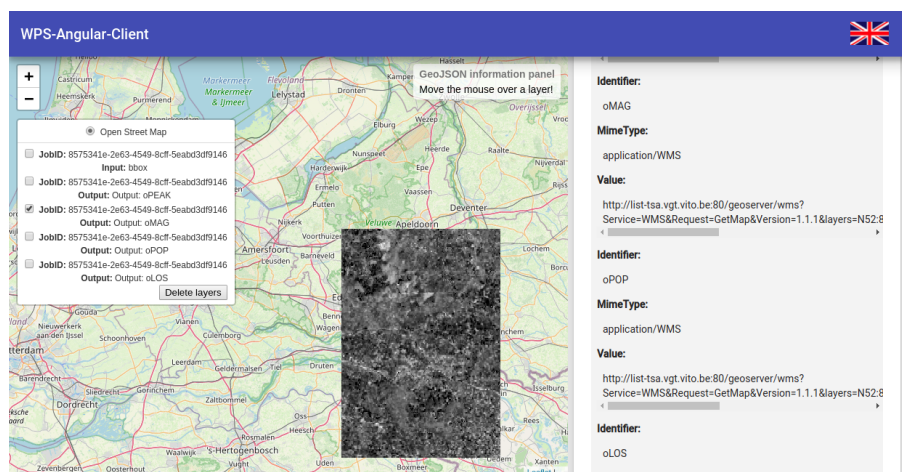


Figure 2: Screenshot of the Web client showing the results of phenology workflow.

## Example: Phenology

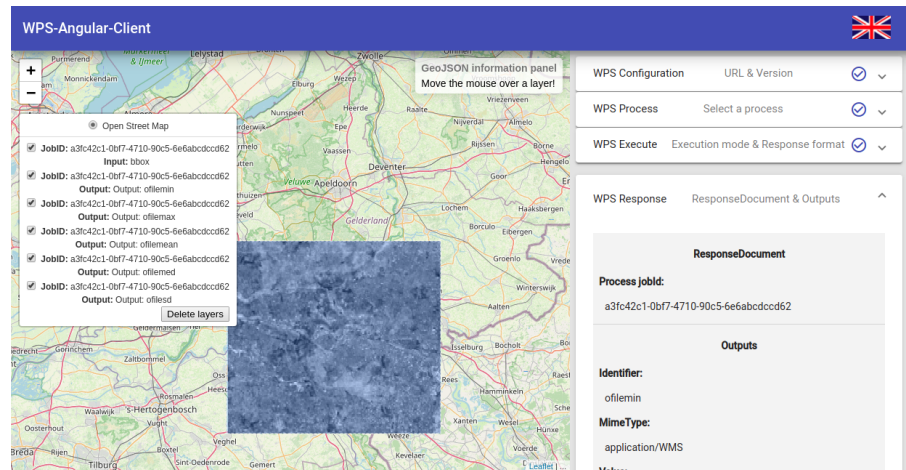


Figure 3: Screenshot of the Web client visualizing an overlay of the process' results.

We have implemented a variety of predefined analysis workflows in the same fashion. These workflows range from summary statistics over linear model fitting, trend analysis and phenological analysis to the automated fitting of autoregressive models. The definition of a temporal window and the spatial extent is common to all workflows. Some also allow for additional parameters, e.g. whether missing data should be filled through an interpolation over time.

The workflow dedicated to phenology derives several characteristics per year from the pixel time series. To achieve this, the remote imagery 10 daily time series is aggregated to monthly maxima and further analysed using the R package [greenbrown](#). The results are a single raster file for each derived property (peak value, magnitude, point of peak and length of season) with one band for each year. Figure 3 illustrates the returned

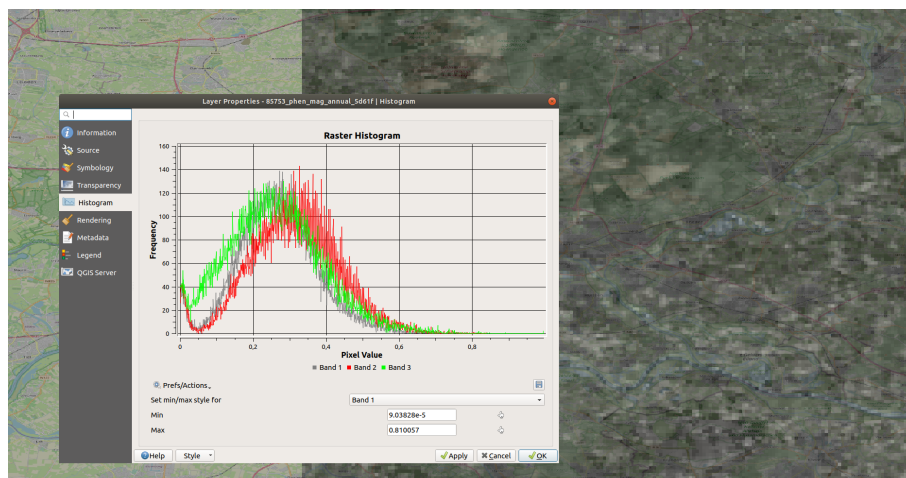


Figure 4: Excerpt of a screenshot of the result of the phenology workflow imported into QGIS. The histogram depicts the distribution of the different annual magnitudes for the study period 2016 (grey), 2017 (red) and 2018 (green).

magnitudes of such an analysis in the Web Client. In Figure 4, the same results have been imported through the WCS interface into QGIS. In the screenshot, the histograms of the three bands depict the distribution of the magnitudes for the three years studied 2016, 2017 and 2018.

## Technical design

The analytical functional core is the R package EOTSA. The functions available in the package are wrapped in a series of services. Using the Web client, the user selects and parametrises a process and sends a WPS execution request to the Web server encapsulating the WPS and its component WPS4R. The WPS4R component allows for the extraction of WPS processes directly from annotated R scripts and makes them available in the WPS component and provides the execution of the R scripts.

Pre-defined workflows or single functions are wrapped in dedicated EOTSA scripts that are sent to the Spark cluster for execution. Running on the master node, the executed workflow or function reads the data from geotiffs according to the temporal and spatial bounding boxes defined in the WPS process. After the data is split into chunks, it is distributed in the cluster and returned upon completion of the analysis. The results are merged and imported into the GeoServer instance, which provides the results to the Web client or any WCS or WMS client. One can also use other entry points to the system and extend the EOTSA toolbox, such as RStudio on a virtual machine within the VITO infrastructure to perform an analysis on the Spark cluster using parallel computation.

Processes exposed through the WPS allow for the integration of the EOTSA toolbox into tools following the OGC WPS specification, such as e.g. the Automatic Service Builder (ASB). Therefore, not only predefined workflows, but all R functions in the EOTSA package have been wrapped as dedicated WPS processes. This enables the user to customise an analysis workflow interactively via drag&drop. This custom workflow can include other processes from other WPS as well. Further details of the entire set-up are summarized in Figure 5.

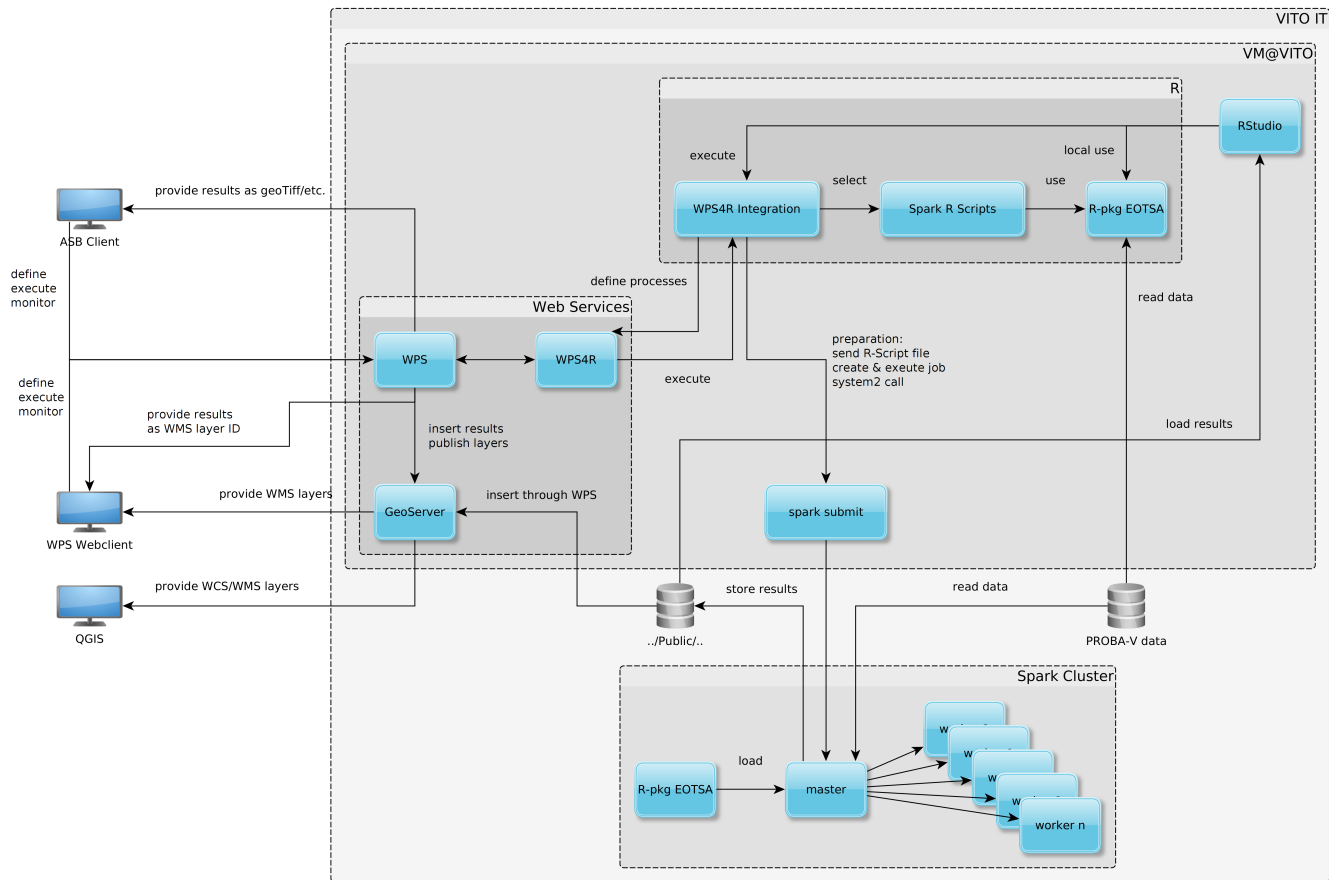


Figure 5: Schema of the interconnected components for the EOTSA toolbox deployment at VITO.

## Conclusion

The EOTSA toolbox eases the access to the data and considerably lowers the data exchange rate between the MEP and users' machines. Re-using a single functional unit across different application scenarios (from local to WPS processes integrated in foreign workflows) achieves a high level of workflow standardisation and harmonisation at a minimal level of analysis maintenance. The R package can not only be used in this dedicated set-up, but can also serve as analysis toolbox for remote sensing data provided as RasterStacks in R. This project proves a strong commitment towards collaborative science.