Food and Agriculture Organization of the United Nations

SEPAL
System for earth observation data access, processing & analysis for land monitoring

Peatlands restoration monitoring working session manual

Bogor, Indonesia
19 – 23 August

Contact:
remi.dannunzio@fao.org
erik.lindquist@fao.org
yelena.finegold@fao.org

Table of Contents

Peatlands restoration monitoring working session manual 1
Table of Contents 2
Objective 4
Use of the terminal in SEPAL 4
    Using the terminal to start an instance 4
    What type of instance do I need ? 5
    Keeping an instance active 6
    6
    Shut down a existing instance 7
    What can I do in the Terminal ? 7
Connecting your Google account to SEPAL 8
EXERCISE 1: Soil moisture mapping 10
Soil moisture content mapping 10
    Installing PYSMM 11
    Processing soil moisture 14
Optional Exercise 1 16
Downloading the soil moisture maps from Google Drive to your SEPAL account 17
    Creating a common folder with shared soil moisture maps 18
Post-processing and analyzing the soil moisture maps 18
Download/Upload data from/to SEPAL to/from your computer 19
    Connect FileZilla to your SEPAL account 19
    Using FileZilla 20
EXERCISE 2: BFAST Explorer 22
Part 1: Open SEPAL 22
    A. Open SEPAL and Login 22
Part 2: Connect SEPAL to your Google account 23
    A. Check if your SEPAL account is already connected to your Google account 23
Part 3: BFAST explorer 24
    A. Open BFAST explorer 24
    B. Map Tab 25
    C. Analysis Tab 27
    D. Tutorial Tab 30
EXERCISE 3: Downloading time series 31
Part 1: Open SEPAL 32
    A. Open SEPAL and Login 32
    B. Connect SEPAL to Google account 32
Part 2: Access and download time series data 32
    A. Search for data 32
    B. Downloaded data in SEPAL 33
EXERCISE 4: Time series analysis 35
Part 1: Open SEPAL 36
    A. Open SEPAL and Login 36
Part 2: Start a session in SEPAL through the terminal 36
    A. Shut down any existing instance 36
    B. Using the terminal to start an instance 36

Part 3: Time series analysis app.....38
A. Open the time series application..... 38
B. Download the test data set..... 39
C. Run time series analysis over your area of interest..... 40

Objective

The objective of this exercise is to create a time stack of soil moisture maps using free and open tools.

This user manual will guide the workshop participants through soil moisture content mapping tools and time series analysis of vegetation indices.

Use of the terminal in SEPAL

One of the ways SEPAL provides access to cloud computing infrastructure is through Amazon Web Services (AWS) cloud computing.
In SEPAL you can select from a variety of cloud computers of various sizes.

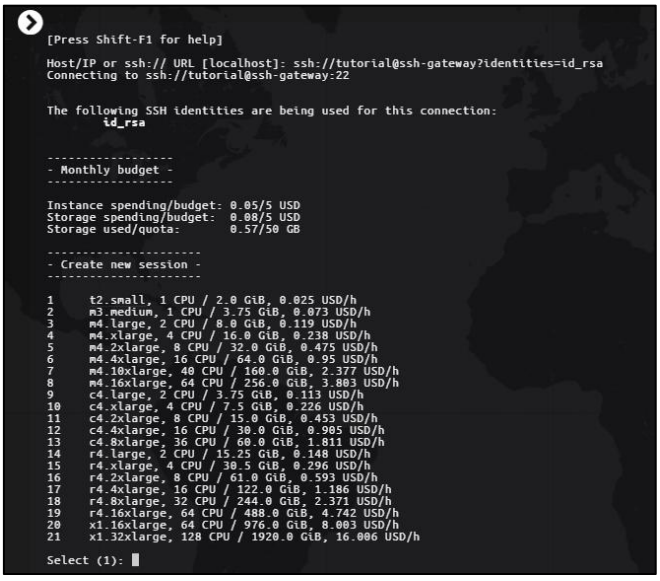
These cloud computers are referred to as **instances** and when they are switched on, you are running a **session**. An instance will automatically start if you launch an application in the processing options. The instance that automatically starts is always the smallest available one, the t2small, which has 1 GB of CPU and 2 GB of RAM power, for a small price of 0.03 USD per hour.

If the session is not active for more than 15 minutes (i.e. you are not using any of the processing tools nor are you actively running anything in the terminal), the instance will be automatically shutdown.

Using the terminal to start an instance

Go to the terminal

If you don't have any instance running you should see the option to start a new session and the list of instances (see image below)



What type of instance do I need ?

It is good practice to adapt the type of instance to your needs.

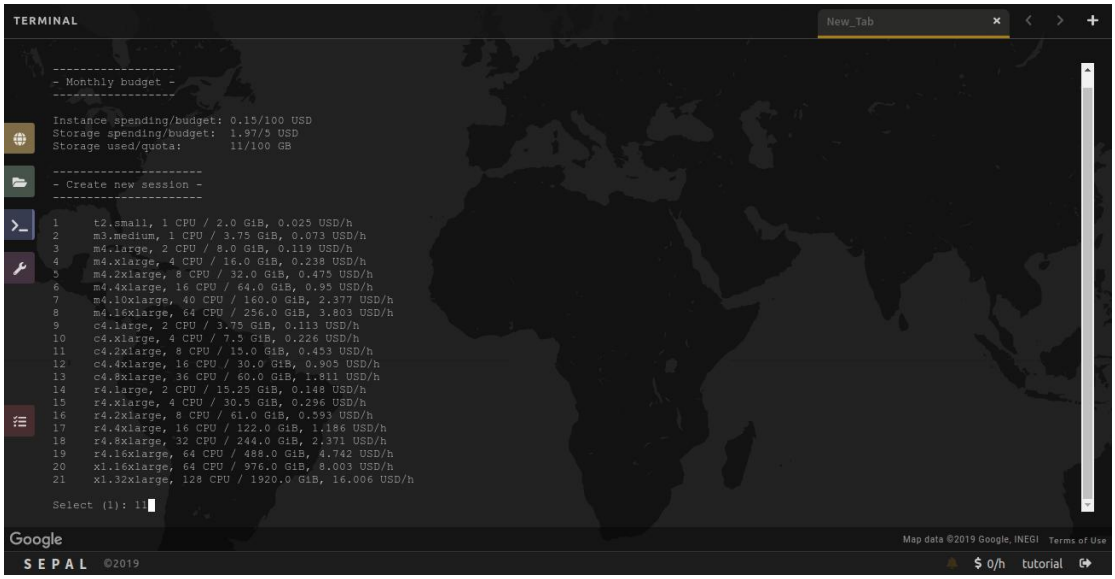
If you are running the soil moisture content processes in Jupyter notebook, you do not need much processing power because the bulk of the processing is sent to Google Earth Engine. You can use a t2.small, instance #1 or m3.medium, instance #2.

If you want to run calculations in R with big objects (national scale raster), you will require a lot of RAM (at least 16GB, instance #4)

The time series analysis application is computationally intensive. You will need an instance with more CPUs and less memory because the algorithm is parallelized among many CPUs and does not require a huge amount of memory. Running a larger instance saves you time and money in your SEPAL account, depending on the size of the area. If the area is large, processing can take days to complete if using a small instance.

For running the time series analysis test data set, you can first try selecting the c4.xlarge. If you are planning on running the time series analysis application on your own area of interest, consider using a larger instance, such as c4.4xlarge or c4.8xlarge

To select the instance you need to type the number corresponding to the instance you would like to start. For example if you are starting the m3.medium, type **2** in the terminal and then hit the ENTER key on the keyboard. The command line only works by typing text in the terminal.



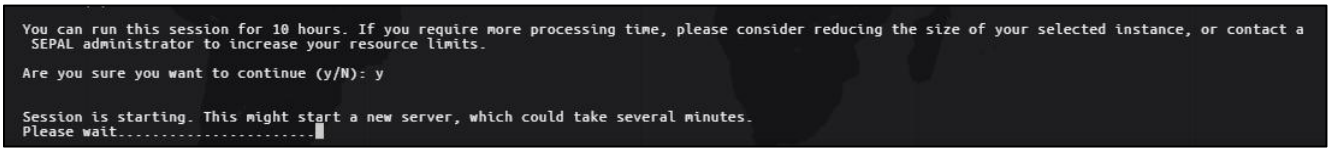
You will then see text about the amount of time you can leave your session running according to your user resources.

i. If your user resources are limited you will be asked if you would like to proceed with starting this instance.

ii. If you would like to proceed, type **y** (which stands for yes) in the terminal. Otherwise, if you would like to select another instance type **n** (which stands for no)**.**

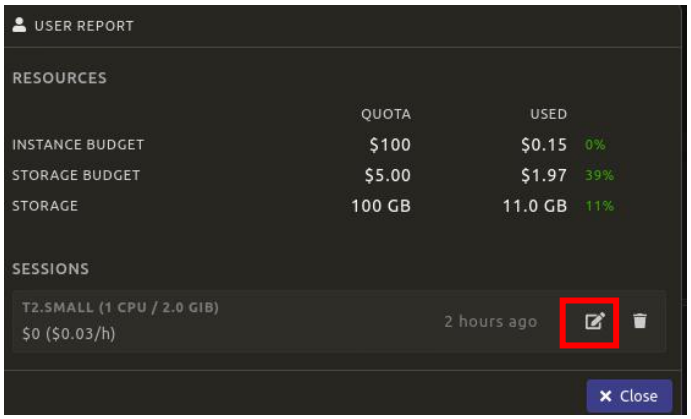
iii. If you have plenty of resources you will not see this text.

It can take a minute to start an instance. You will see text that says 'Please wait….'

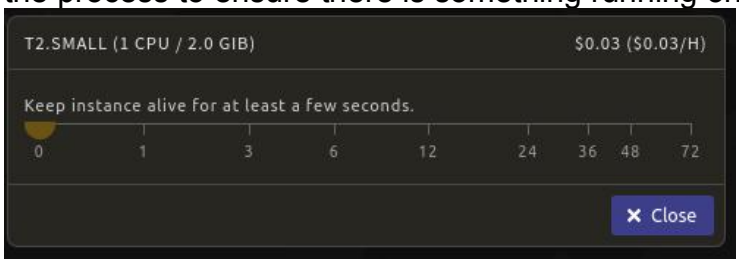


When the session has completed loading you will see your username@numbersandletters$

Keeping an instance active



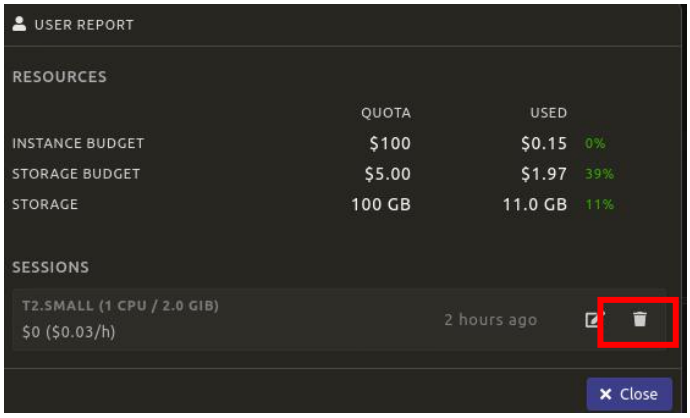
Click on the edit symbol
Use the slider to choose the amount of time to keep the instance alive. You can keep the instance alive for up to 72 hours. Only use this feature if you are using a process that you know needs to an active instance, such as the time series analysis application or soil moisture mapping. Make sure to check on the process to ensure there is something running on the instance.



Shut down a existing instance

You can manually shutdown an instance in your user resources, which you can access by clicking on the instance budget.

Click the next to the running instance to shutdown the existing instance.



What can I do in the Terminal ?

SEPAL is running under a Linux server with the latest long term release Ubuntu operating system and the corresponding Bash Shell. You can run all the basic Linux commands in the terminal. In addition, several geospatial processing libraries are available such as GDAL/OGR (www.gdal.org) or the Orfeo Toolbox (www.orfeo-toolbox.org) and you can run processing commands from the terminal.
You can also clone, update and push git repositories directly in the terminal as illustrated in the below examples.
Clone a repository:

```
git clone https://github.com/yfinegold/ws_idn_20190819
```

Go to a repository:

```
cd   ~/ws_idn_20190819/
```

Update a repository:

```
git pull
```

If there are conflicts between your local GitHub repository and the updated repository , then you can discard the local changes you made in your local repository.
Be careful using this! It will throw away any of your local changes in the files from the repository. This includes any of the scripts but not the data folder.

```
git reset --hard
```

Then don't forget to git pull after resetting. Git pull will update your repository to the latest version.

Update a repository:

```
git pull
```

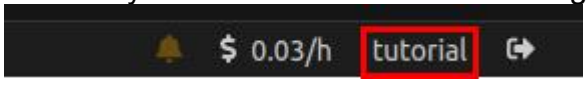
Connecting your Google account to SEPAL

Some of the core functionality of SEPAL utilizes Google Earth Engine (GEE). To take full advantage of SEPAL you need to connect your Google account to your SEPAL account.

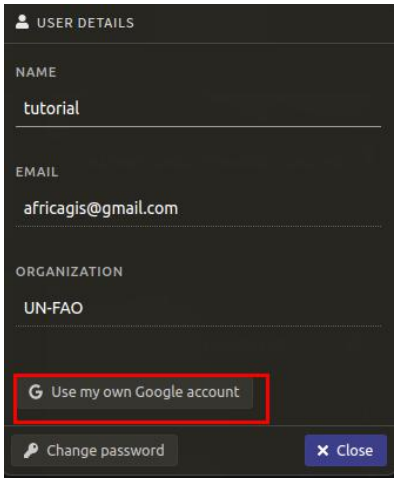
First, make sure you have a Google account and it is whitelisted in GEE.
If you do not have a GEE enabled account, sign up here:
https://signup.earthengine.google.com/

After you get an email confirming your access to Google Earth Engine you can connect your SEPAL account to Google.

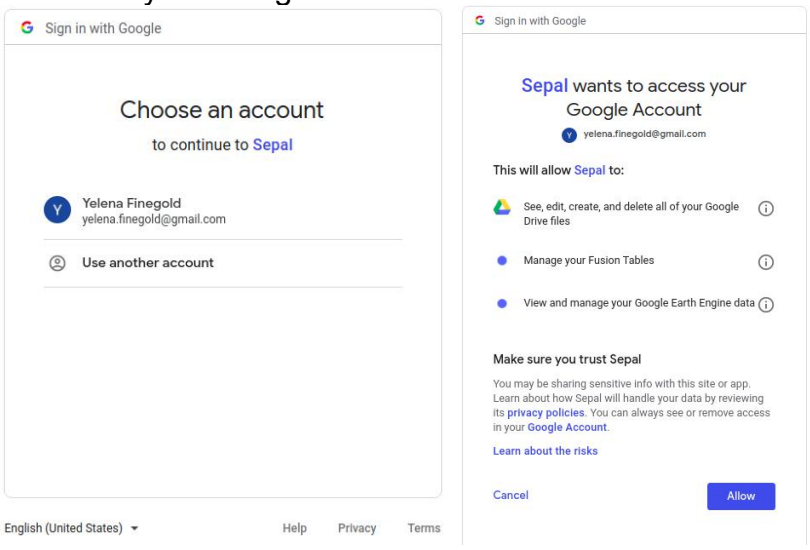
Click on your user name in the bottom right corner



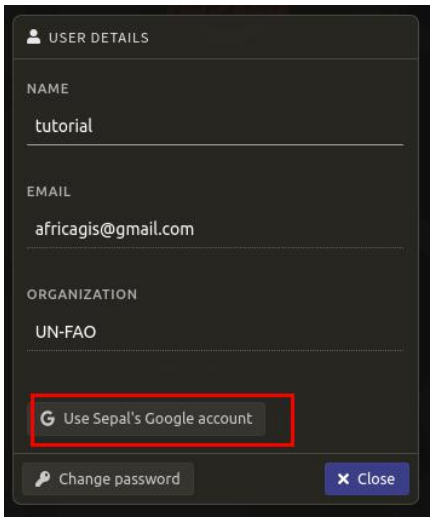
If you see 'Use my own Google account', your Google account is not yet connected to SEPAL. In the following steps you will connect your account. First, click on 'Use my own Google account'. This will redirect you to a Google sign in page.



Select the Google account that is already approved for use of GEE. Sign into your account with your Google password. Then click on ‘Allow’ to grant SEPAL access to your Google account.



After clicking ‘Allow’, you will be automatically redirected to the SEPAL page. You can check your Google account connection by clicking on your user name again. It should now say ‘Use SEPAL’s Google account’. Do not click on this button, it will disconnect the connection between your personal SEPAL account and your personal Google account.



EXERCISE 1: Soil moisture mapping

Soil moisture content mapping

The soil moisture content module creates soil moisture maps using the PYSMM process developed by, Felix Greifeneder, read more about PYSMM here: https://pysmm.readthedocs.io/en/latest/.

The estimation of soil moisture is based on a Support-Vector-Regression machine learning approach. The model training was performed based on in-situ data from the International Soil Moisture Network (ISMN). PYSMM all processing steps for spatial and temporal mapping of surface soil moisture are fully executed online on GEE - and then the data sets are transferred to your SEPAL account.

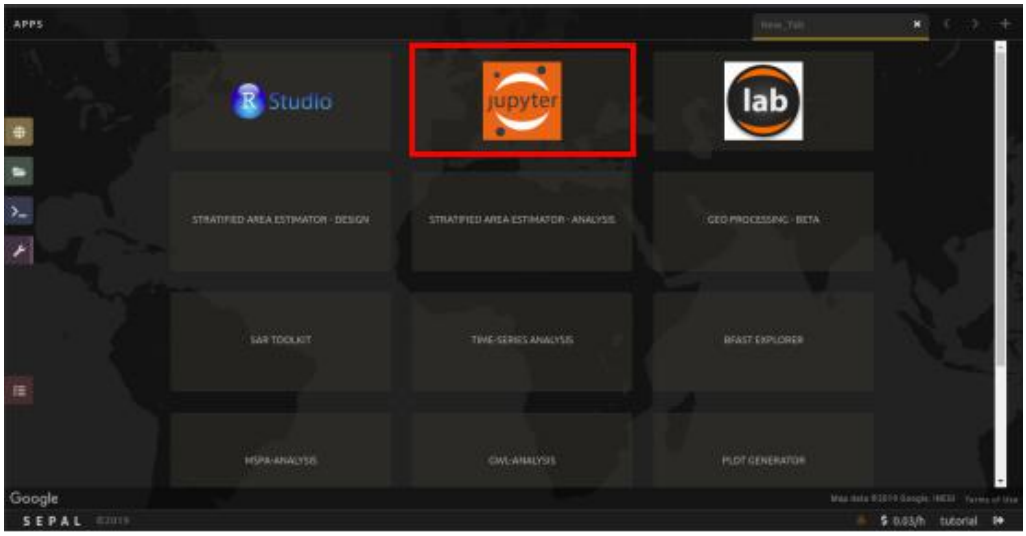
Jupyter notebooks interfaces have been created to processing the soil moisture maps for an area of interest. The area needs to be uploaded as a Google Earth Engine asset. To upload an asset follow the instructions here: https://developers.google.com/earth-engine/importing

The PYSMM modules are part of the ws_idn_20190819 folder which was downloaded using the git clone command in the command line. It will be necessary to update this folder periodically using the instructions in the Terminal section of this manual.

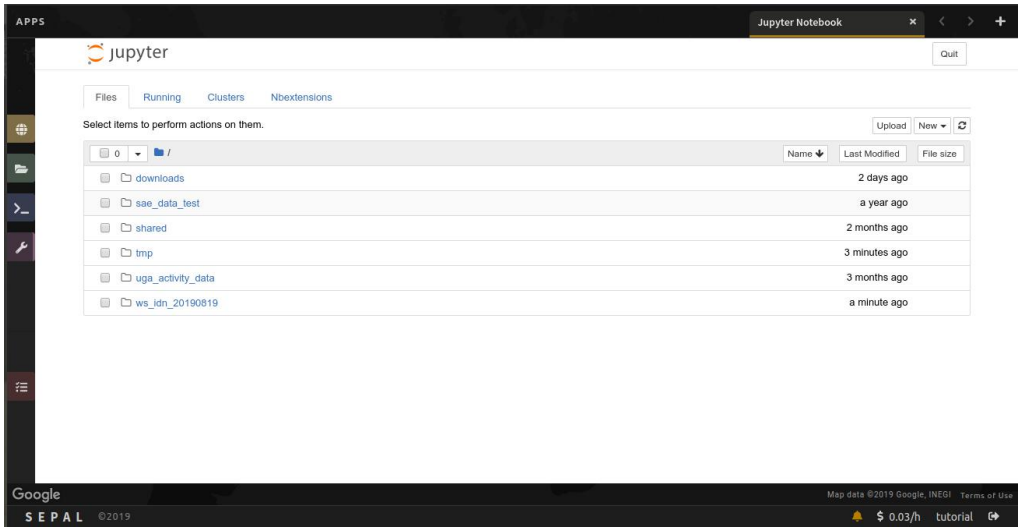
If you would like to learn more information about soil moisture mapping, this site is informative: https://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9569

For more information about SAR in general, this very comprehensive handbook on how to use SAR data for forest monitoring and biomass estimation: https://gis1.servirglobal.net/TrainingMaterials/SAR/SARHB_FullRes.pdf

To access the modules click on the processing tab in SEPAL. Then click on Jupyter notebooks.

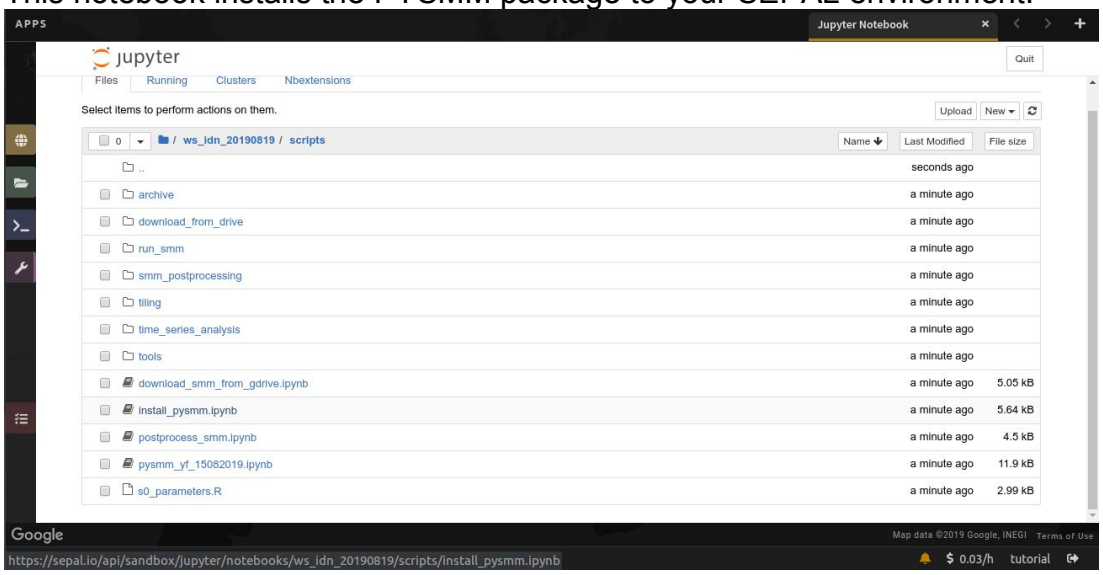


Then navigate to the ws_idn_20190819 folder by clicking on it. Then click on the scripts folder.

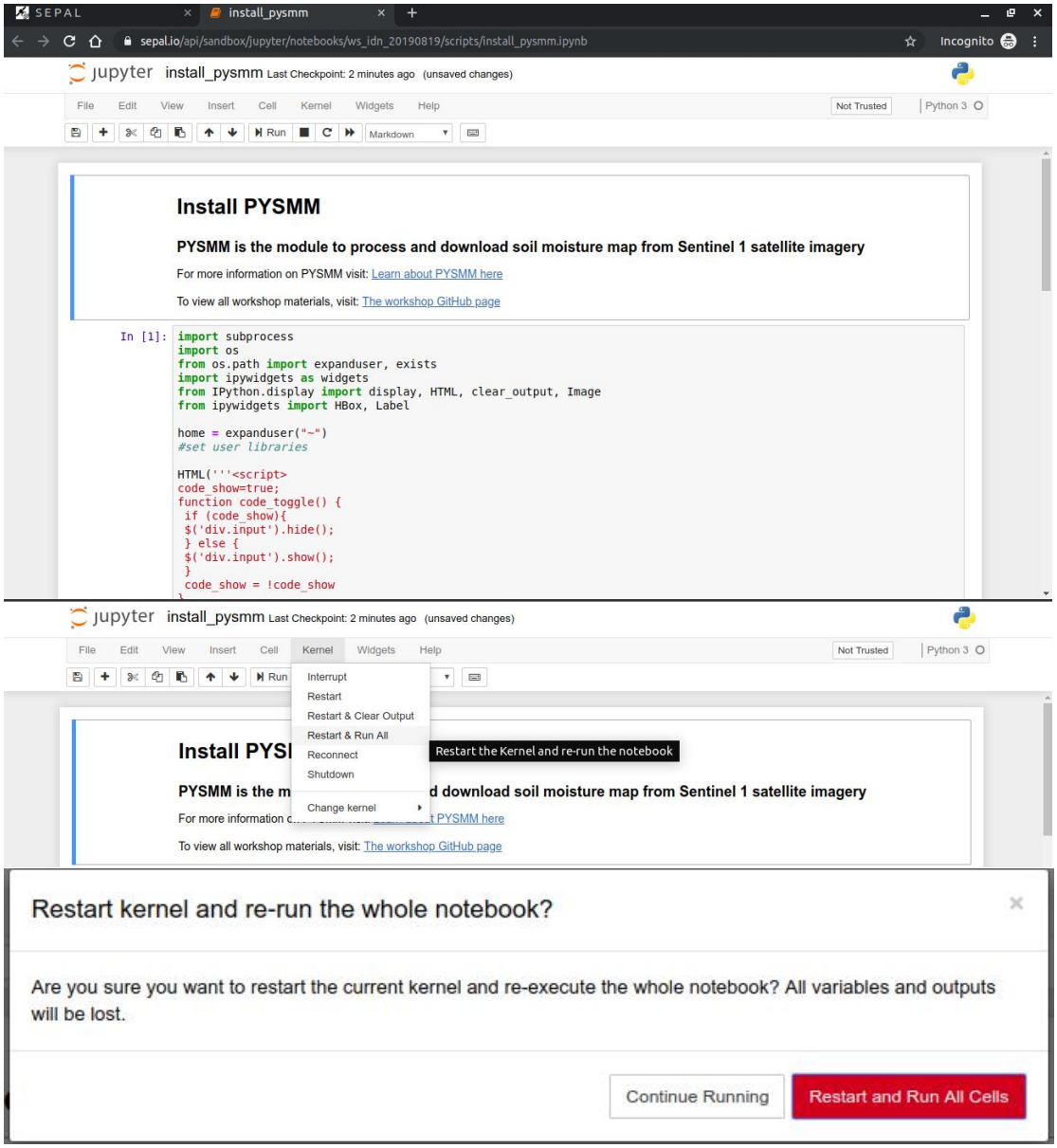


Installing PYSMM

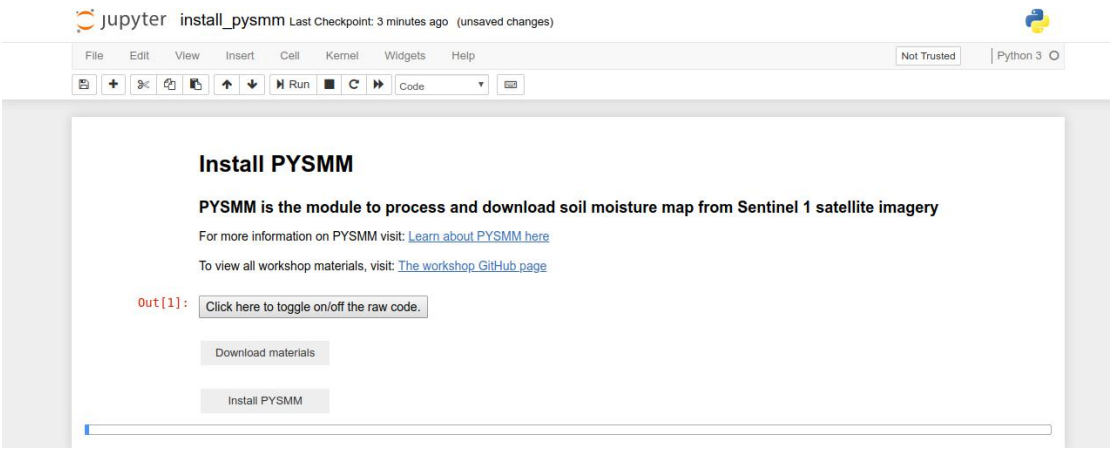
Inside the scripts folder there are python notebooks, which end with the .ipynb extension. Click on the install_pysmm.ipynb to open the first python notebook. This notebook installs the PYSMM package to your SEPAL environment.



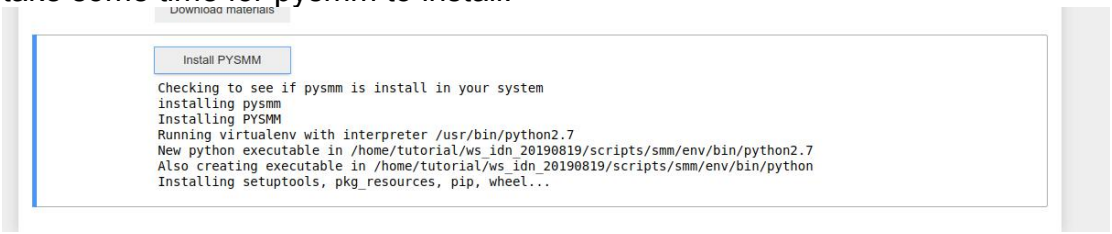
The jupyter notebook will open in a new tab. First you need to restart the kernel and run all. To go this click on ‘Kernel’ and then click on ‘Restart & Run All’. Wait a moment for the notebook to initialize.



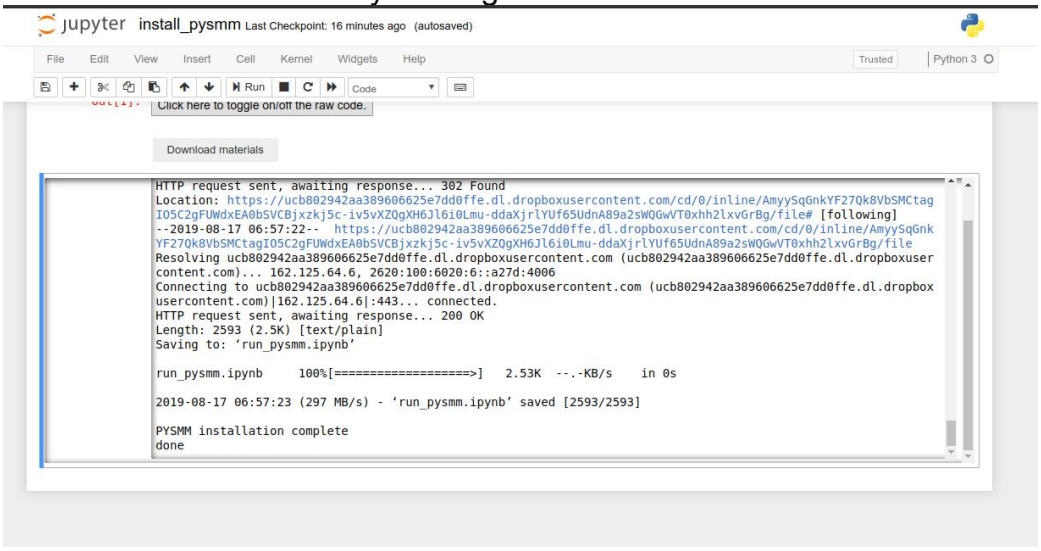
Once the notebook is initialized, the code should not be visible and only the user interface remains. There is the option to toggle the code on and off. If you are interested in seeing the code you can toggle the raw code, although this is not necessary.
To install pysmm click on the ‘Install PYSMM’ button. You will only need to install pysmm one time.



You will see the installation text appear with updates about the process. It will take some time for pysmm to install.

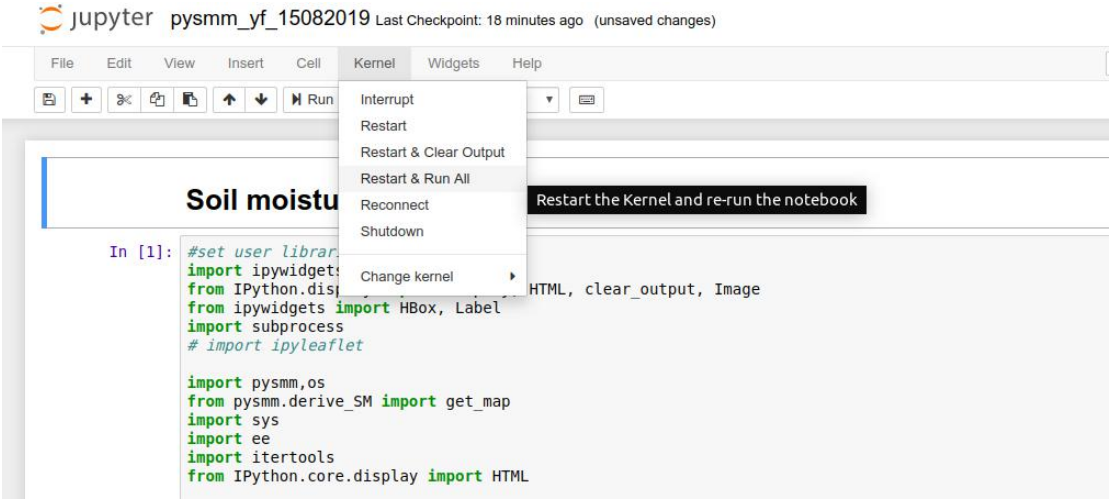


When the installation is completed, scroll down to the bottom of the installation updates and it will say, ‘done’. When you complete the installation you can shut down the notebook by clicking file->close and halt.

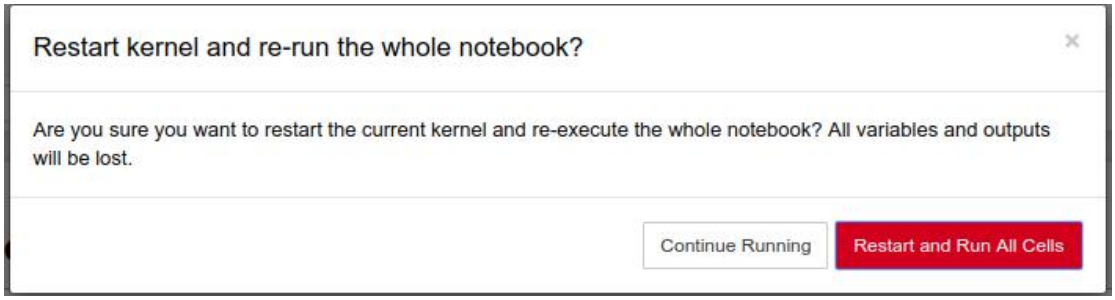


Processing soil moisture

The python notebook to process soil moisture is called run_pysmm_ui.ipynb. Open the notebook from jupyter notebooks and reset the interface by selecting 'Kernel'->'Restart & Run All'

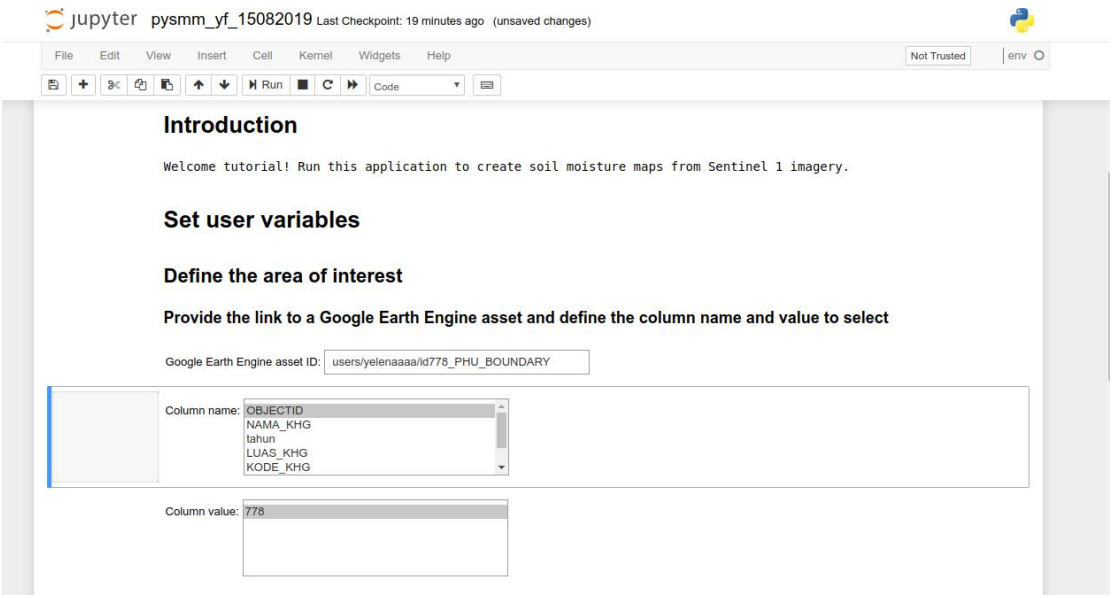


Click on Restart and Run All Cells button

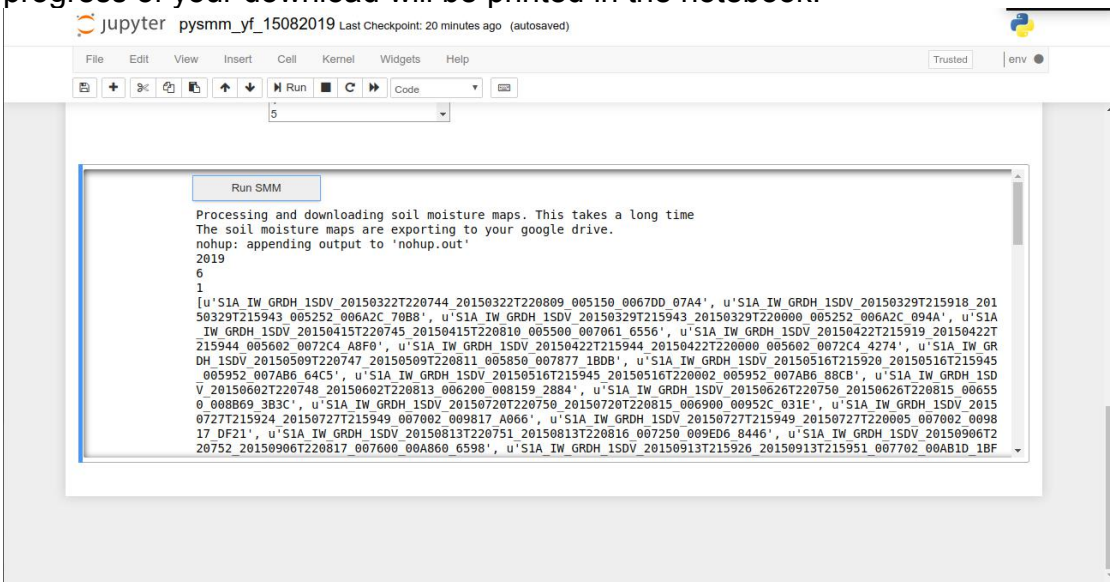


The area of interest is specified using a Google Earth Engine table asset. A Google Earth Engine table asset is vector data, such as a shapefile, uploaded to your GEE account. You can change the area of interest by pasting your GEE asset id into the text box. If you change the asset id, click on the white space left of the column name box and type CTRL and ENTER at the same time to read your area of interest file and load the custom column names. The area of interest is filtered by rows in the attribute table. To select the column value, first select the row you want to chose the value from. Then click left of the column value box and type CTRL and ENTER at the same time to read your area of interest file, selected column names and load the values from that column.

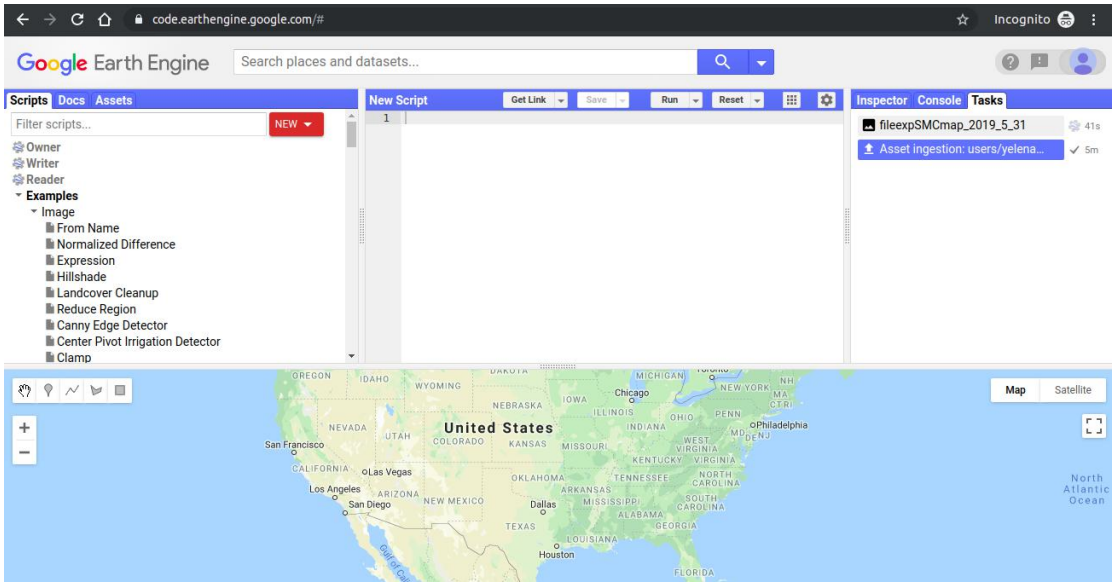
Next select the date parameters for downloading the soil moisture maps.



After all the parameters have been customized you can run the soil moisture module. This will send the command to Google Earth Engine to run the PYSMM process. Your data will be downloaded into your Google drive. The progress of your download will be printed in the notebook.



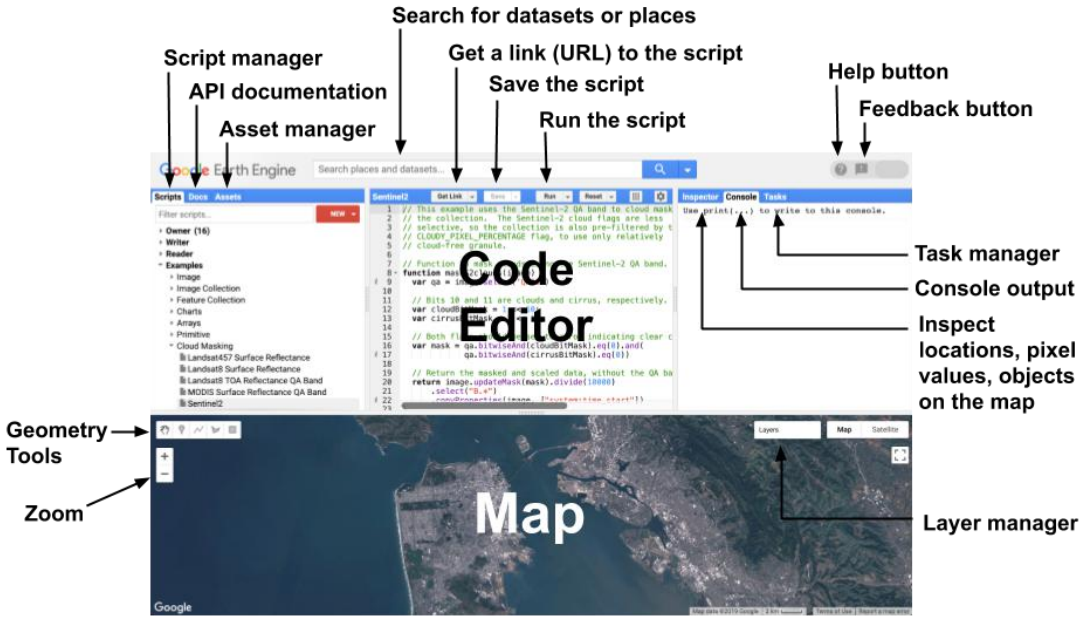
Another way to check on the status of your download is to go to https://code.earthengine.google.com. Click on the 'Tasks' tab in the section on the right. You should see the process running with the spinning gear. When the download completes you will see a blue check mark. Check periodically on your download to make sure all the dates specified are being downloaded.



Optional Exercise 1

While waiting for the soil moisture maps to download lets explore the availability of Sentinel 1 imagery over Indonesia. Click on this link to see the number of acquisitions of Sentinel 1 globally: https://code.earthengine.google.com/6c919eaa51cb77507e373af8eca3fbc7

Here are the components of Google Earth Engine:



For more information about using Google Earth Engine, check out: https://developers.google.com/earth-engine/playground

Guiding questions

Choose an area that you work in. How many Sentinel-1 images are available over your area of interest?

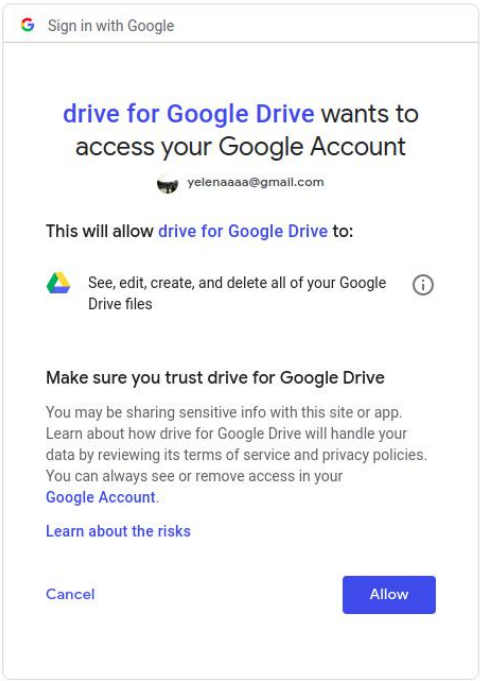
How many Sentinel-1 images are available over a location in Germany?

Downloading the soil moisture maps from Google Drive to your SEPAL account

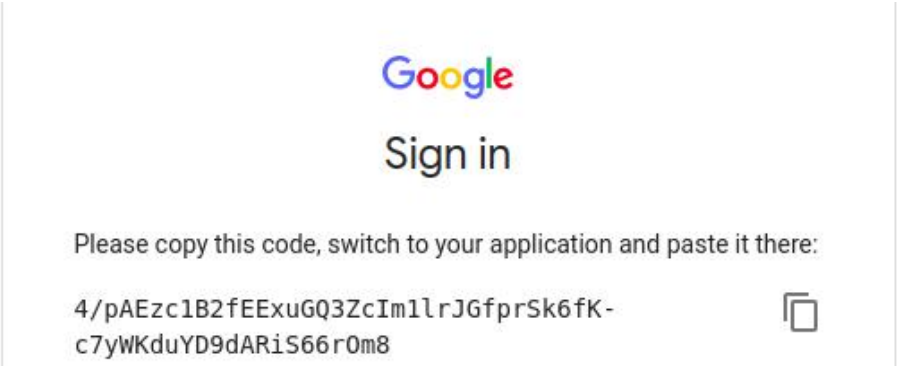
After the download of the time series of soil moisture content maps in complete, check your Google Drive, you will see newly downloaded SMCmaps in your home folder. To copy these images between your cloud storage in Google and your cloud storage in SEPAL is very easy.

In Jupyter notebooks open up the download_smm_from_gdrive.ipynb Jupyter notebook in the scripts folder.
Click on Kernel>Restart and Run All as in the previous notebooks.

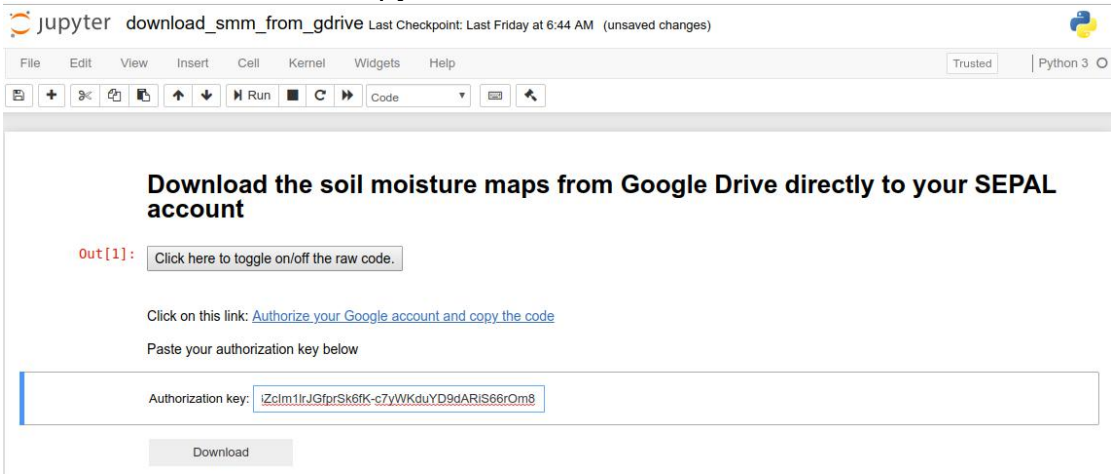
First click on the link to authenticate your Google account.This must be the same Google account you connected to SEPAL and where you soil moistre maps are downloaded.
Click allow.



Copy the authorization code by clicking on the button to the left of the code or by using copy.



Paste this code into the Jupyter notebook.



Then click Download to start the data transfer from Google drive to your SEPAL account.

After the data download completes you can use tools available in SEPAL to process and analyze these soil moisture maps.

Creating a common folder with shared soil moisture maps

After the download completes move your soil moisture maps in Google drive to the shared Google drive folder. Create a folder with your SEPAL username and inside that folder create a folder describing the area of interest you have downloaded. For the example data set, call it ‘example’. Move the data into the folder.
The shared Google Drive can be found here:
https://drive.google.com/drive/folders/1PJRpjJZjydHZIUyi0rF5SnrWiXbtLQf9?usp=sharing

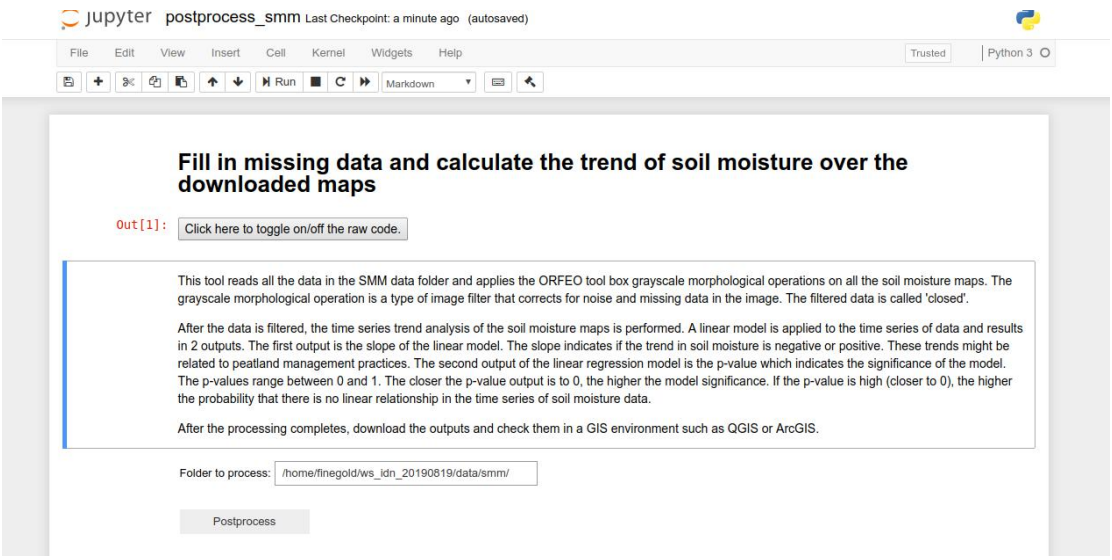
Post-processing and analyzing the soil moisture maps

After your download of the soil moisture content is complete we can apply a robust methodology for image filtering to fill no data gaps. Due to speckle in the Sentinel-1 imagery the soil moisture maps contain some noise and no-data

values which are corrected for in some extent using grayscale morphological operation from ORFEO toolbox, a free and open source image processing tool. To read more about the parameterization of the Orfeo toolbox tool, read: https://www.orfeo-toolbox.org/CookBook/Applications/app_GrayScaleMorphologicalOperation.html.

After filtering the image and filling in no data gaps in all the downloaded soil moisture maps in the specified folder, this Jupyter notebook applies a linear regression over the time series of soil moisture map to help understand if there is an increasing or decreasing trend of soil moisture content and the significance of this trend. A linear model is applied to the time series of data and results in 2 outputs. The first output is the slope of the linear model. The slope indicates if the trend in soil moisture is negative or positive. These trends might be related to peatland management practices. The second output of the linear regression model is the p-value which indicates the significance of the model. The p-values range between 0 and 1. The closer the p-value output is to 0, the higher the model significance. If the p-value is high (closer to 0), the higher the probability that there is no linear relationship in the time series of soil moisture data.

To run the post-processing module specify the folder you where your soil moisture maps are located and then click the 'Postprocess button'.



After the post-processing completes download the data from SEPAL to your personal computer for analysis and interpretation of the results.

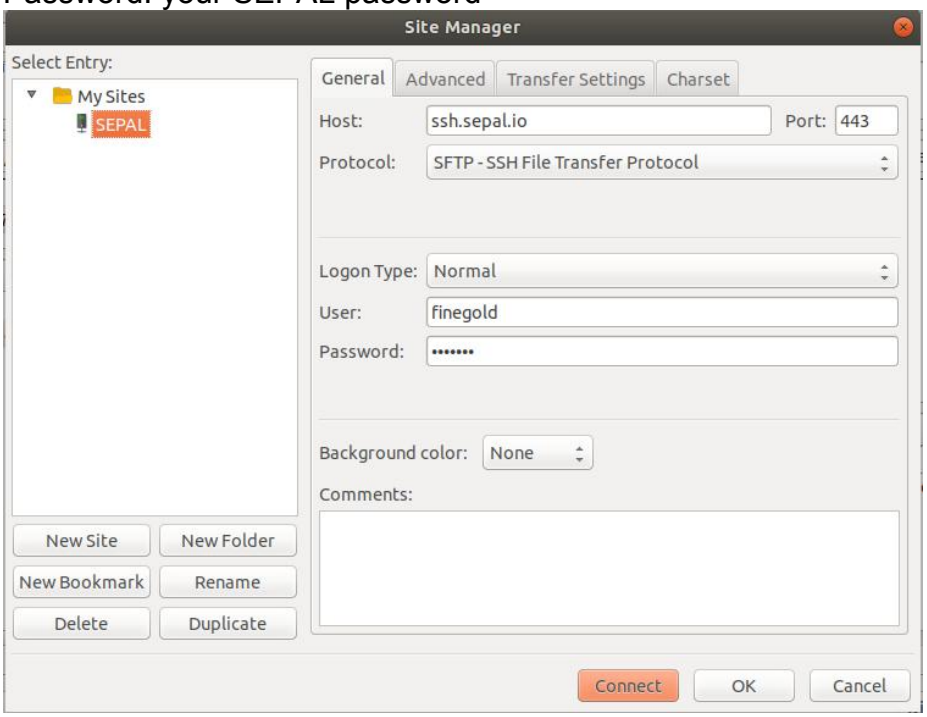
Download/Upload data from/to SEPAL to/from your computer

Connect FileZilla to your SEPAL account

Download FileZilla from this link

Accessing files in SEPAL is easy using FileZilla. To use FileZilla, open the application and connect to the SEPAL server by selecting Menu File --> Site Manager in the menu tab.

In the site manager pop-up click on the 'New Site' button. Use the screenshot below as a guide for filling in the form:
Host: ssh.sepal.io
Port: 443
Protocol: SFTP – SSH File Transfer Protocol
Logon Type: Normal
User: your SEPAL username
Password: your SEPAL password



Click Connect and your SEPAL drive will appear in the right panel, parallel to the panel on the left with the files on your computer.

Using FileZilla

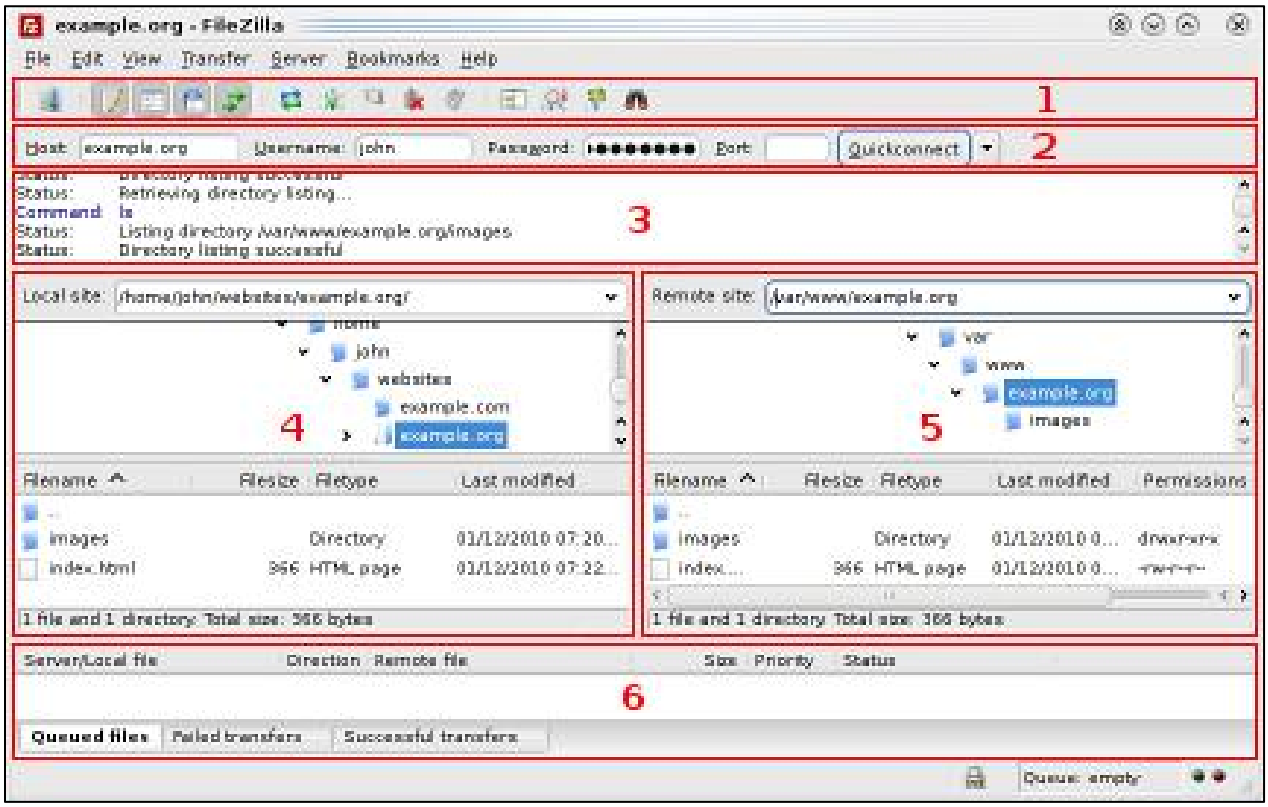
Here is a quick introduction from the FileZilla online tutorial corresponding to the image below:

i. the toolbar (1)
ii. quick connect bar (2),
iii. the message log (3) displays transfer and connection related messages. Below, you can find the file listings.
iv. The left column (local pane, 4) displays the local files and directories, i.e. the stuff on the PC you're using FileZilla on. Make sure you are

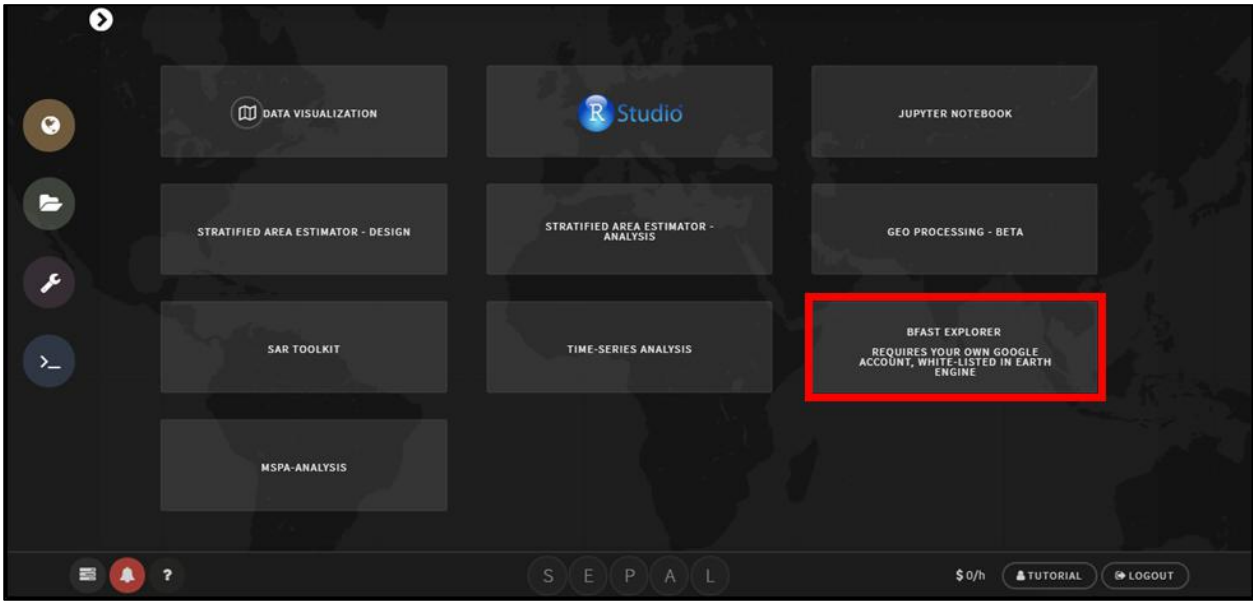
located in the right folder. Ideally, you should have a folder on your computer that will store all data related to this process, point to that folder

v. The right column (server pane, 5) displays the files and directories on the server you are connected to. Both columns have a directory tree at the top and a detailed listing of the currently selected directory's contents at the bottom. You can easily navigate either of the trees and lists by clicking around like in any other file manager.

vi. At the bottom of the window, the transfer queue (6) lists the to-be-transferred and already transferred files.



EXERCISE 2: BFAST Explorer



Acknowledgement

Exercise 2 is from the tutorial available inside the BFAST explorer application by Alexandre Esteves Almeida

Introduction

Now that you are familiar with the SEPAL interface, we can start exploring the tools for time series analysis. Exercise 2 introduces time series analysis using the pixel-based BFAST explorer.

Objectives

Learn use the BFAST explorer tool

Become familiar with time series concepts

Test the parameterization for BFAST monitor

Prerequisites

SEPAL account

Google Earth Engine access

Part 1: Open SEPAL

A. Open SEPAL and Login

1. If SEPAL is not already open, click the following link to open SEPAL in your browser: https://sepal.io/.

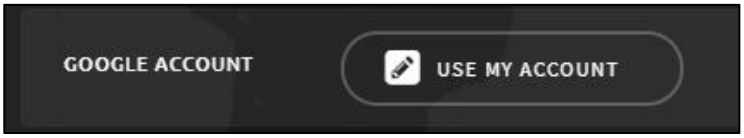
Part 2: Connect SEPAL to your Google account

A. Check if your SEPAL account is already connected to your Google account

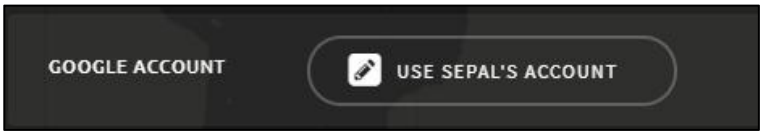
1. Click on your username

2. In the **User** box, there is a row that says **Google account**. There are 2 options here

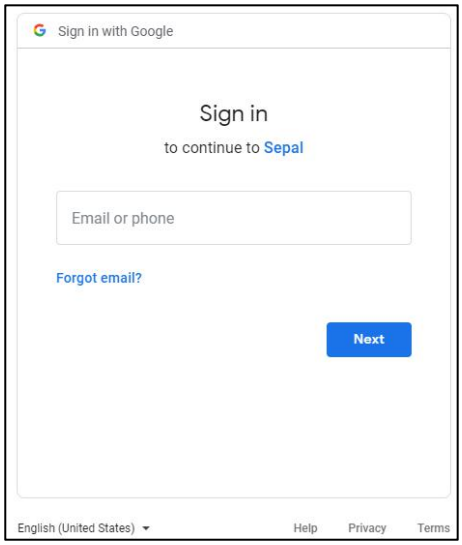
   i. **Use my account**—you need to connect your SEPAL account to google. Proceed with the following steps



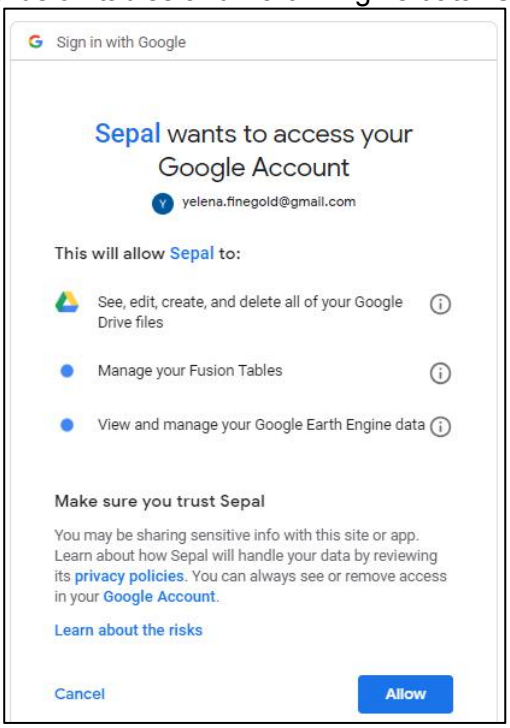
   ii. **Use SEPAL's account**—you are already connected! You can skip the rest of this part and proceed to part 3.



3. Click on **Use my account** from step 2i.
4. You will be redirected to a Google sign in. Please sign in with your google account that is approved for Google Earth Engine access.



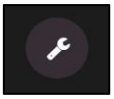
5. After successful sign in, SEPAL will ask for permission to access your google drive, fusion tables and Earth Engine data. Click the **Allow** button to continue.

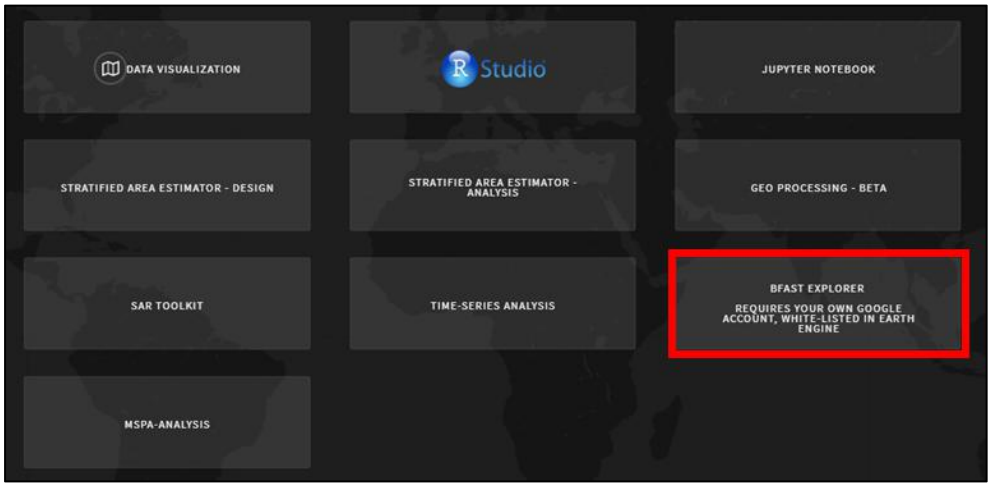


This will take you back to SEPAL and you have connected SEPAL to your Google account.

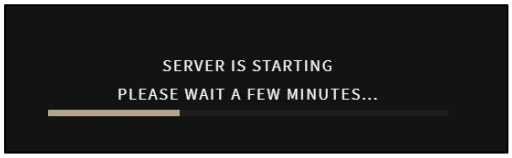
Part 3: BFAST explorer

A. Open BFAST explorer

1. Go to the processing tools 
2. Click on BFAST explorer

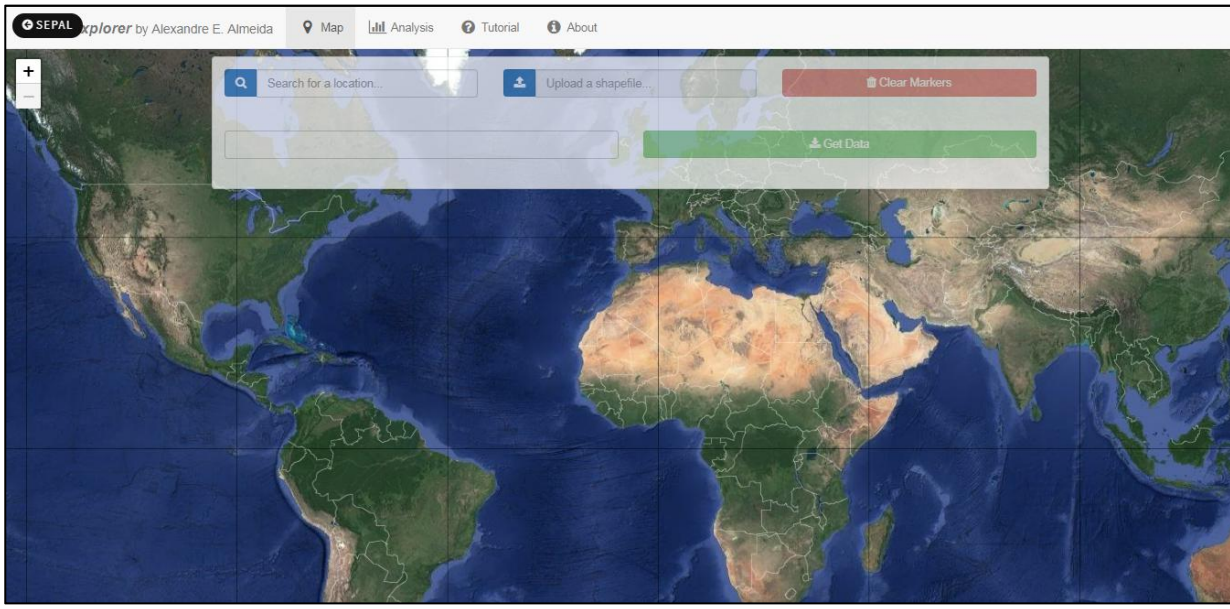


3. If you do not already have an instance running (indicated by the $0/h next to your user name), the smallest instance will automatically start.

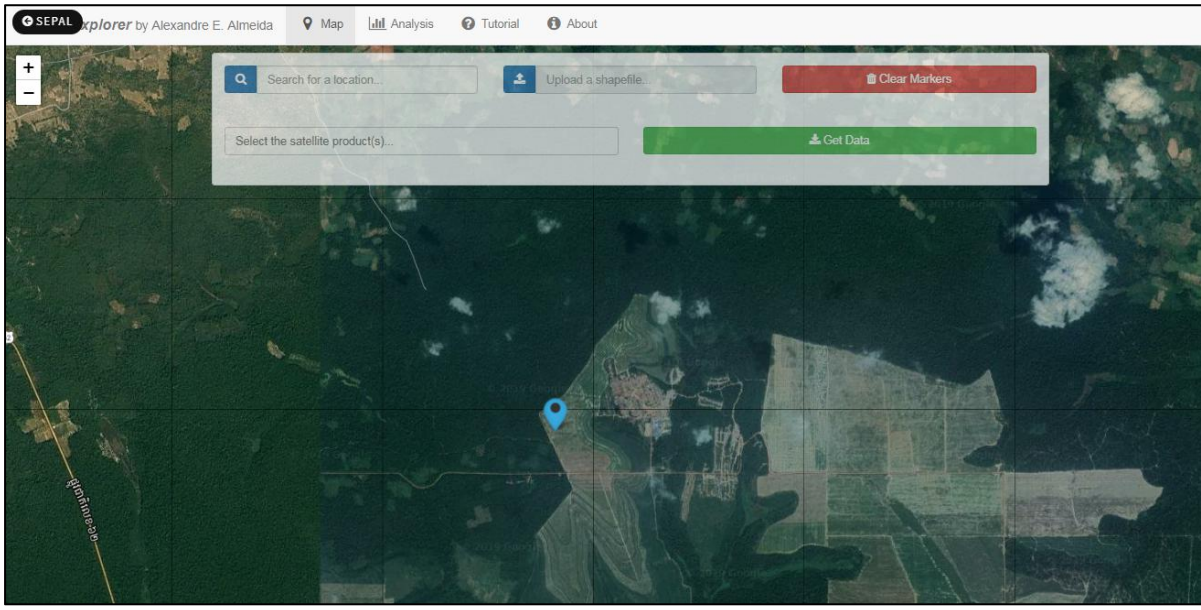


B. Map Tab

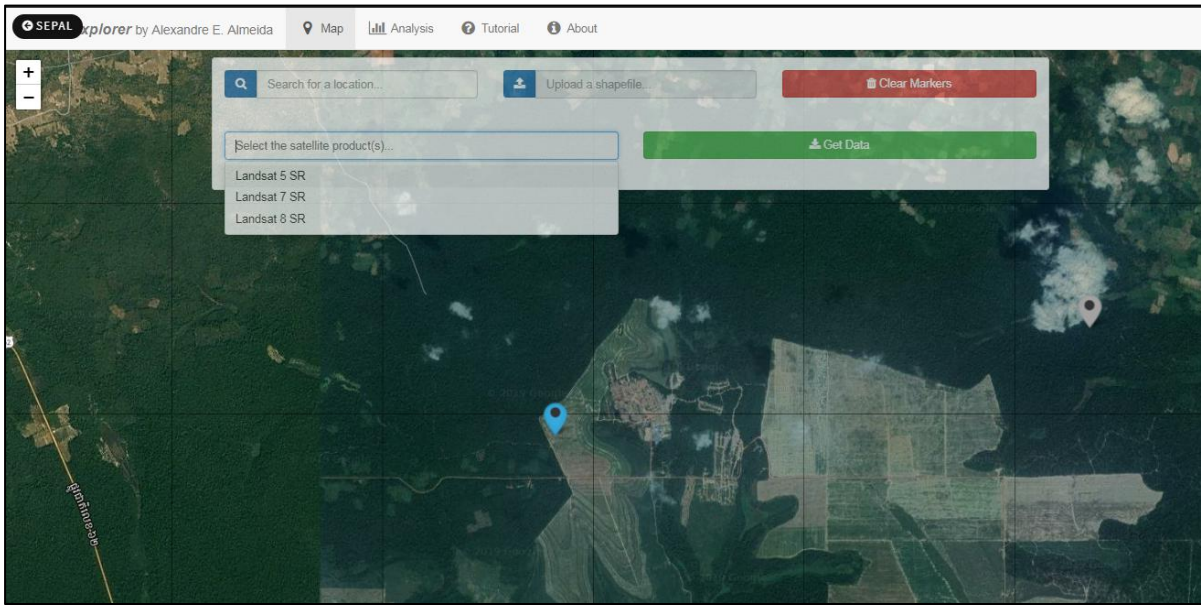
1. This is the starting tab, which we first see when we run the tool. The tab is composed of an interactive map (rendered using Google Maps engine) and a navigation toolbar. Feel free to zoom and pan the map.



2. To *place* a marker, simply click on the map. If we want to, we can also place multiple markers.
3. We may also wish to clear all the placed markers. To do that, click on the **[Clear Markers]** red button on the left side of the toolbar.
4. After that, we need to *select* one of the markers in order download its Landsat pixel data. To do that, simply click on an already placer marker, and it will be highlighted, in blue. Only one marker may be selected at a time.

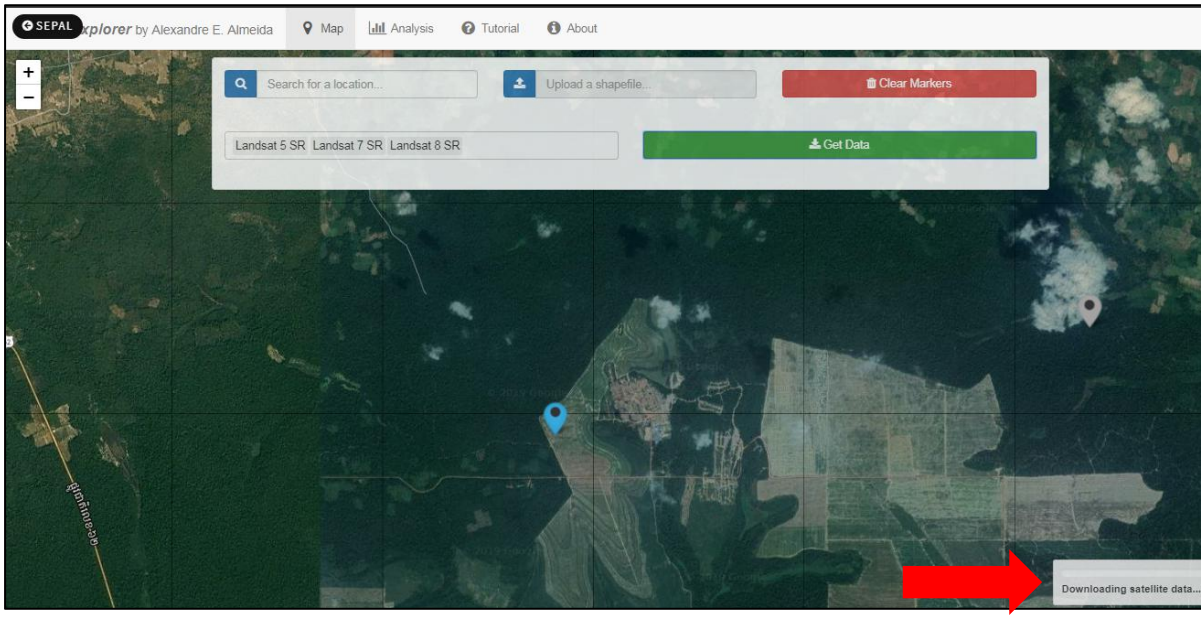


5. By selecting a marker, we can now choose a combination of which satellites to download from using the drop-down menu, located on the bottom of the toolbar. For example, let's choose all the available satellites products: Landsat 5, 7 and 8 SR.

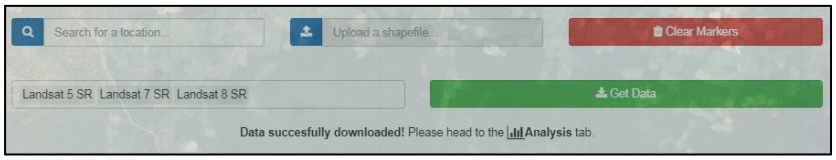


6. Then, we press the **[Get Data]** blue button, located on the right side of the toolbar. By pressing that button, the download will start. We can keep track of the download progress by looking to the lower right corner. All the historical data available are downloaded, which should take less than 10 seconds for the three products selected.

Note: as of the writing of this guide, not all Surface Reflectance data are available from GEE. So, depending on where we place our markers, we may face a message indicating that 'No data available for the chosen satellite(s) and/or region… Please change your query and try again.' *Since we rely heavily on GEE to download the data, there's nothing we can do yet. We're sorry for that.*

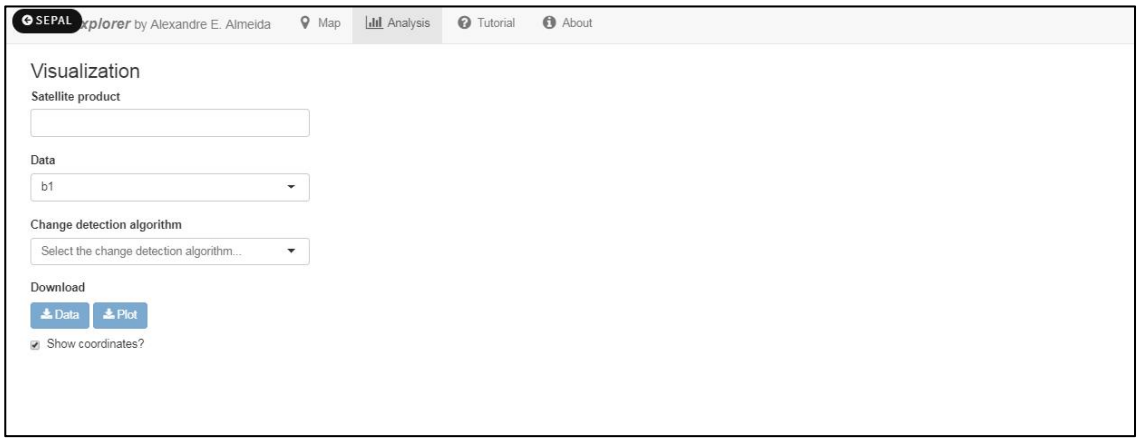


7. If the download is successful, we'll receive a message directing to the **Analysis** tab.



C. Analysis Tab

1. In this tab, we can analyze the downloaded data and, then, locally save the results as files.



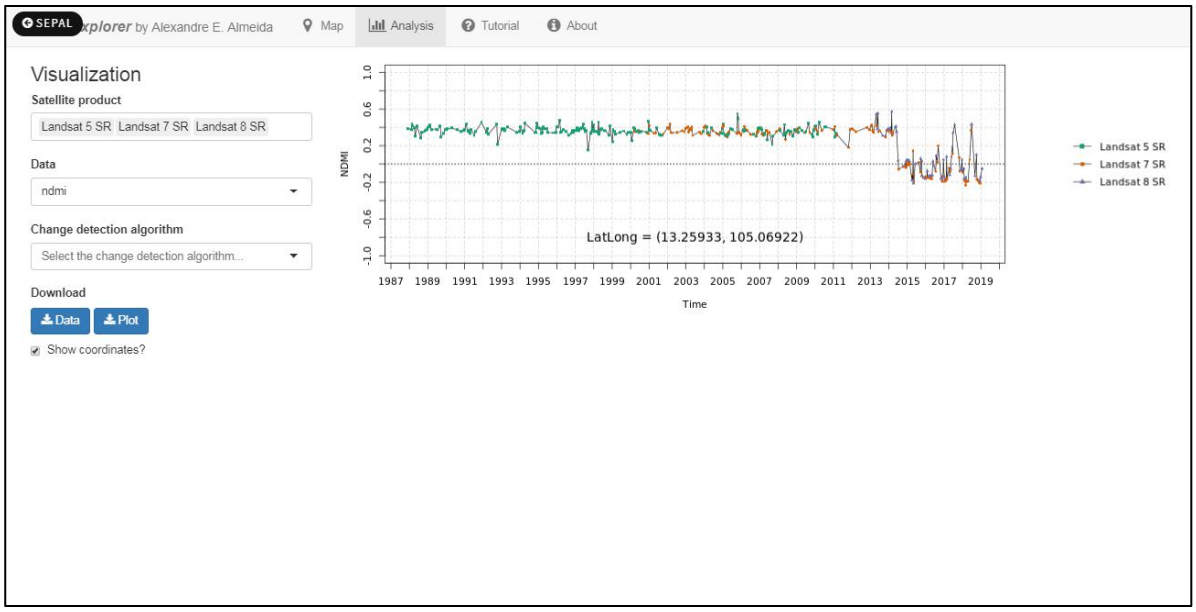
2. First, let's choose which satellite time series date to visualize. Note that, even though we downloaded data from Landsat 5, 7 and 8 SR, we're can still analyze them separately. However, let's proceed by choosing all of them.
    i. As we can see, the time series of the first spectral band (`b1`) is plotted for all satellites. A colored legend distinguishes the different sources.

Note: be careful when comparing spectral bands *data from different satellites, as they may not correspond to the same wavelength range! Read more about this here.*

3. Apart from the spectral bands, there are also four spectral-bands-derived indexes available: NDVI, NDMI, EVI and EVI2. Let's check, for example, the NDMI time series.

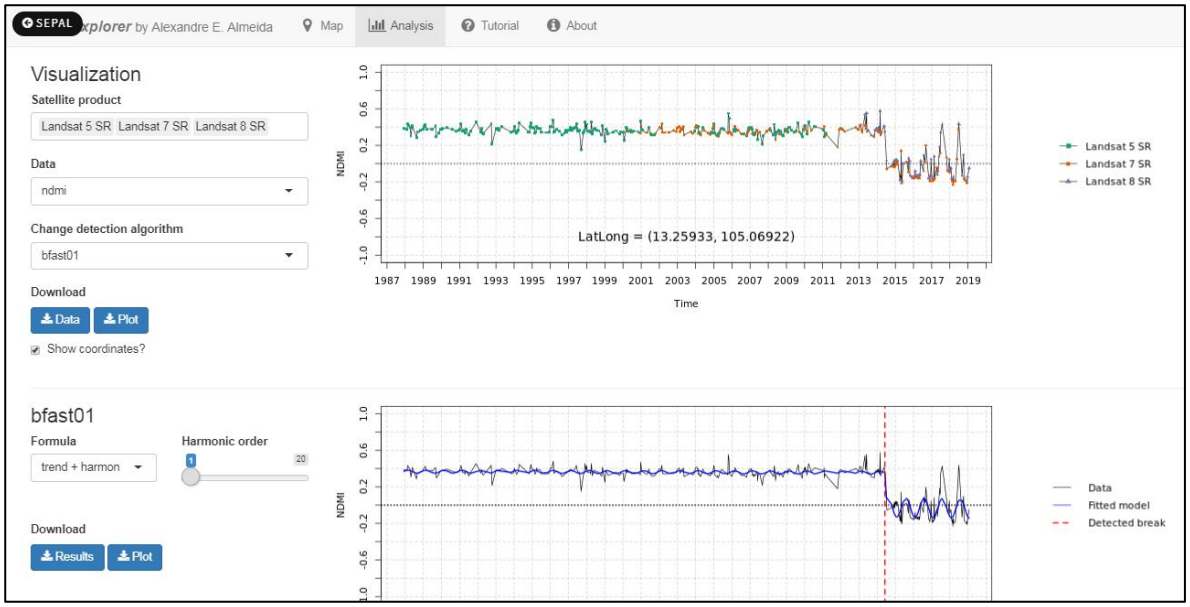
4. If we want to, we can also download *all* the time series data as a file. To do that, press the **[Data]** blue button. All the data will be downloaded as a .CSV, ordered by the acquisition date. Also, an additional column is included, in order to distinguish the satellite sources.

5. We may download the time series plot as an image, by pressing the **[Plot]** blue button. A window will appear offering some raster (.JPEG, .PNG) and a vectorial (.SVG) image output formats.



6. Next, we select the *change detection algorithm*. Three options are available: **bfastmonitor**, **bfast01** and **bfast**. More information about these algorithms can be found here.

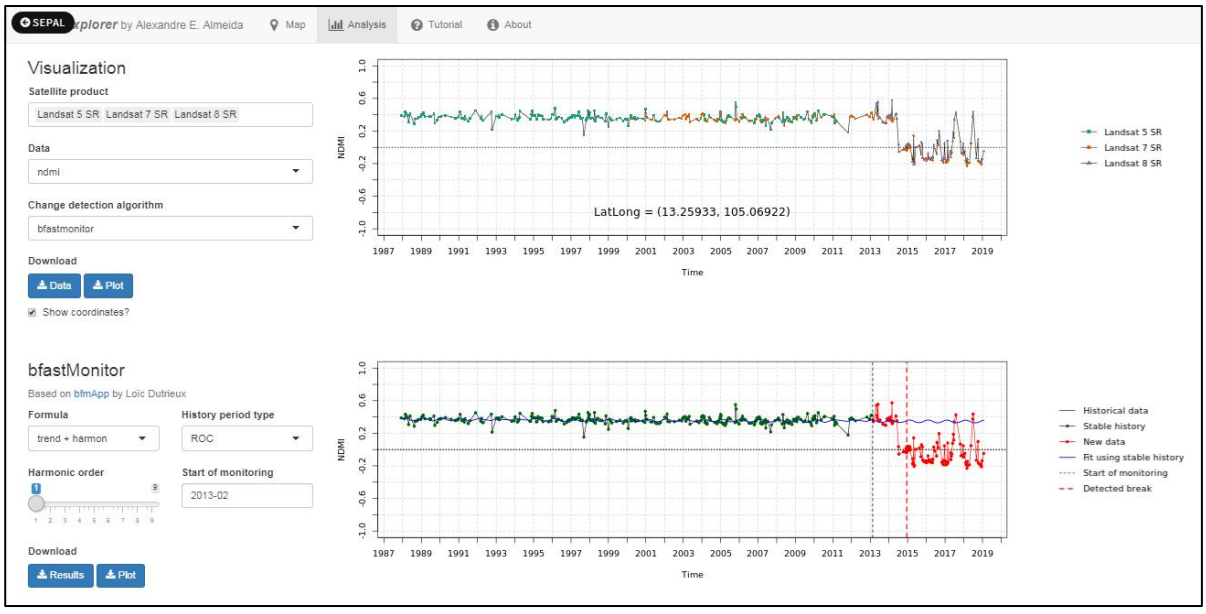
7. By selecting **bfast01**, we can tweak two parameters: formula, and harmonic order. Here, the maximum value of the harmonic order is dynamically set depending on the time series data length and the choice of the formula parameter.



8. By selecting **bfastmonitor**, we are able to tweak four parameters on the left side-bar: formula, history period type, harmonic order, and start of monitoring. These parameters have different impacts on the results, which can be verified on the right side plot. Here, we set the maximum value of the harmonic order to 9 to avoid some problems.

9. Similar to the time series, we can also download the *results* of the change detection algorithms as .RDS data files, by clicking on the **[Results]** blue button. If we wish to download the plot, we can press the **[Plot]** blue button.

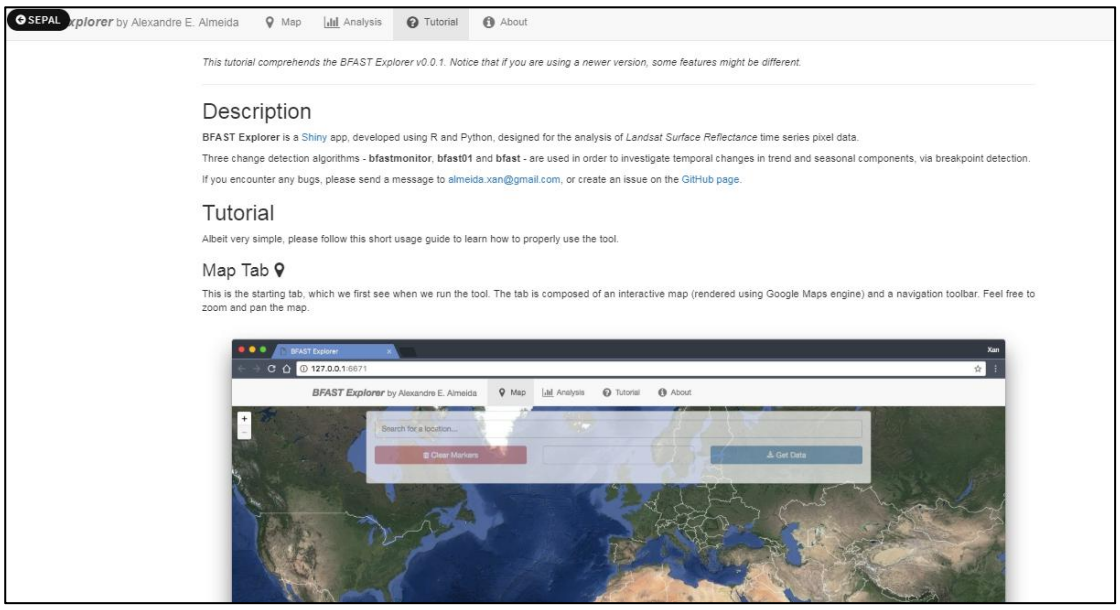
10. For more information on how to load .RDS files on R, please check this link.



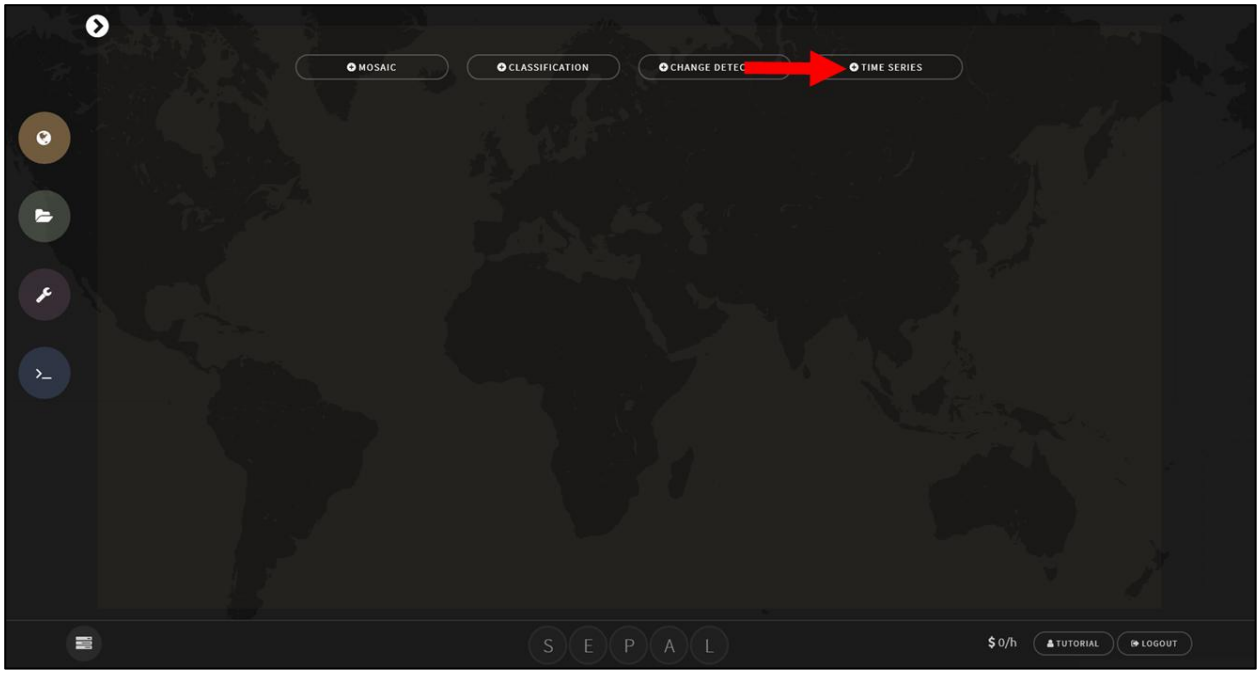
11. Finally, by selecting **bfast**, we may tweak two parameters: h (minimal segment size), and season type. Please note that, since **bfast** can detect multiple breakpoints, it may take a couple of seconds to process, in comparison to the previous two algorithms.

D. Tutorial Tab

1. This tutorial can be viewed inside of the application in the tutorial tab.



EXERCISE 3: Downloading time series



Acknowledgement
Exercise 3 was compiled by Yelena Finegold, Remi d'Annunzio and Erik Lindquist

Introduction
Now that you are familiar with the SEPAL interface, we can start exploring the tools for time series analysis. Exercise 3 enables uses to download a time series of a vegetation index for an area of interest.

Objectives
Download a Landsat time series of a vegetation index for an area of interest

Prerequisites
SEPAL account
Google Earth Engine access

Part 1: Open SEPAL

A. Open SEPAL and Login

1. If SEPAL is not already open, click the following link to open SEPAL in your browser: https://sepal.io/.

B. Connect SEPAL to Google account

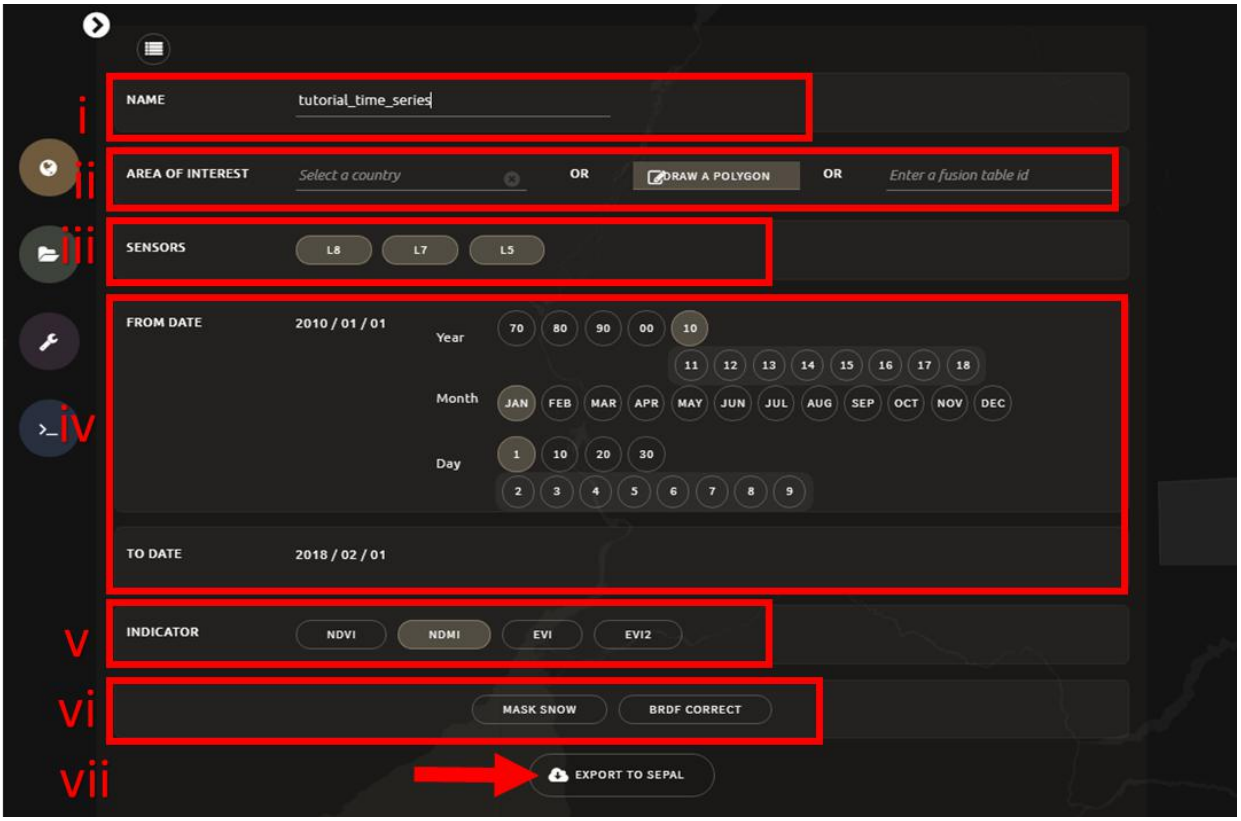
1. Make sure SEPAL is connected to your google account as described in Exercise 2, Part 2

Part 2: Access and download time series data

A. Search for data

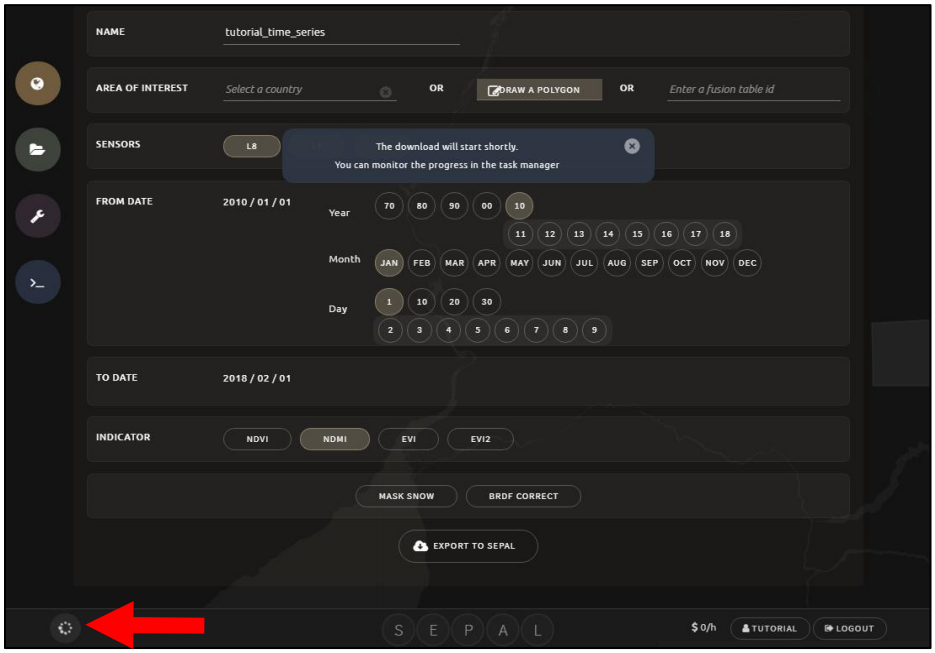
1. Go to the Search module
2. Select the Time Series tab

3. Specify the parameters (see image below):
   i. First give the time series a custom name, this will be the name of the folder in the downloads folder where you will be able to find the downloaded time series data
   ii. Country boundaries, a custom polygon or a fusion table ID can be used as the area of interest. For more information about fusion tables see the Change detection in SEPAL manual.
   iii. Landsat 5, 7 and/or 8 can be included in the time series
   iv. Choose the 'from' and 'to' dates. The time series will start at the *from* date and end at the *to* date
   v. The indicator is the vegetation index that is calculated for each satellite image acquisition.
      (a) NDVI= normalized vegetation index
      (b) NDMI= normalized moisture index
      (c) EVI= enhanced vegetation index
      (d) EVI2= enhanced vegetation index (2 bands)
   vi. Options to mask snow and correct for view and illumination angle effects using BRDF
   vii. Final step is to export the time series stack to SEPAL



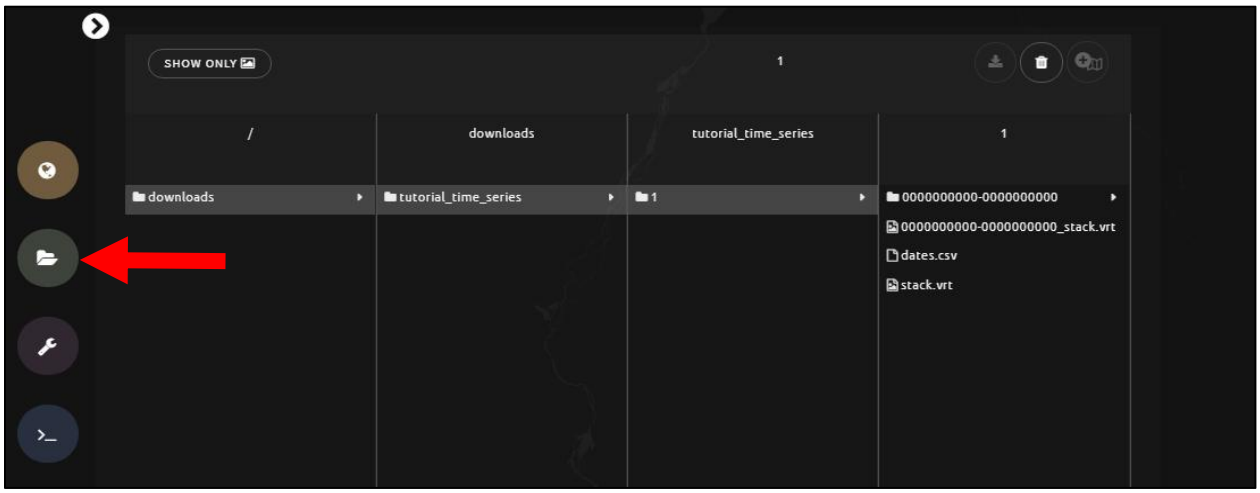
4. The time series tab creates a stack of images for the chosen indicator and time span for the area of interest. In the result each band in the image represents a unique date

5. Once the download is initiated you can monitor the progress of the download by clicking on the spinning wheel in the bottom left corner



B. Downloaded data in SEPAL

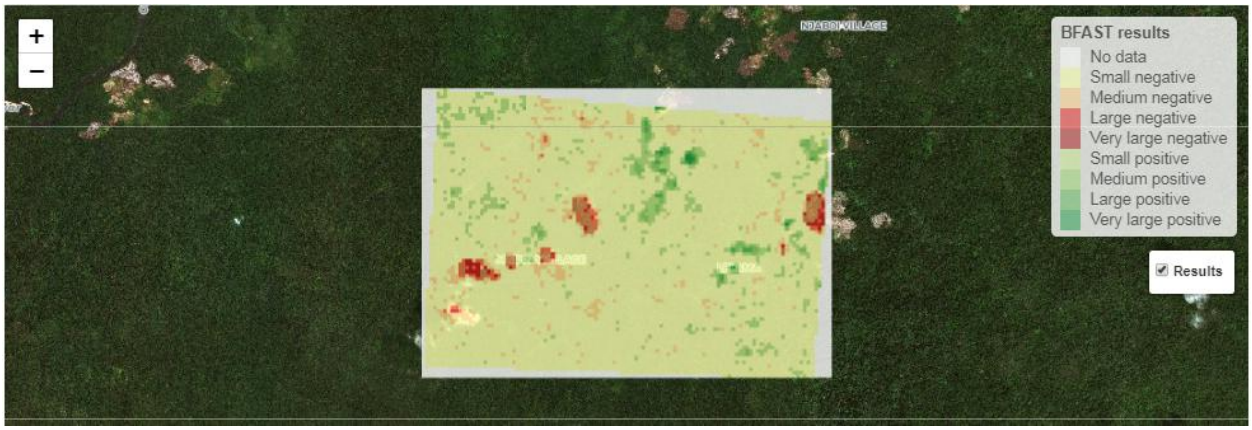
1. The time series stack will download into the download folder in your SEPAL account. Click on the folder icon to see the files.
2. When the download is complete the time series stack is saved as a .vrt file in the downloads folder in a folder that has the same name specified in the download parameters (i)



3. The two main outputs are stack.vrt and dates.csv
   i. **stack.vrt** stores the vegetation index for each date in the bands
   ii. **dates.csv** stores the date corresponding to each band

The data is ready for time series analysis

EXERCISE 4: Time series analysis



Acknowledgement

Exercise 4 was compiled by Yelena Finegold, Remi d'Annunzio and Erik Lindquist

Introduction

Now that you are familiar with the SEPAL interface, had an introduction to time series analysis per pixel and downloaded a time series stack of a vegetation index over your area of interest you can now process this data using the BFAST spatial algorithm.

Objectives

Run BFAST over your area of interest

Prerequisites

SEPAL account
Google Earth Engine access (in order to follow exercise 3 to download your time series data)
Enough budget in your SEPAL account to run the analysis (recommended at least 10 USD processing time)

Part 1: Open SEPAL

A. Open SEPAL and Login

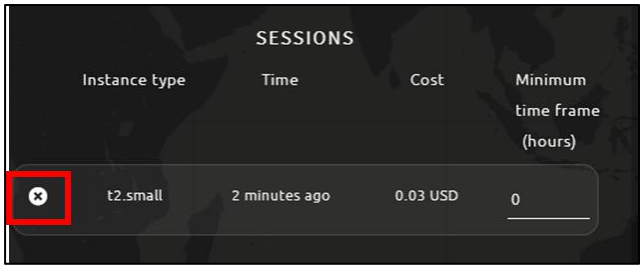
1. If SEPAL is not already open, click the following link to open SEPAL in your browser: https://sepal.io/.

Part 2: Start a session in SEPAL through the terminal

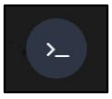
A. Shut down any existing instance

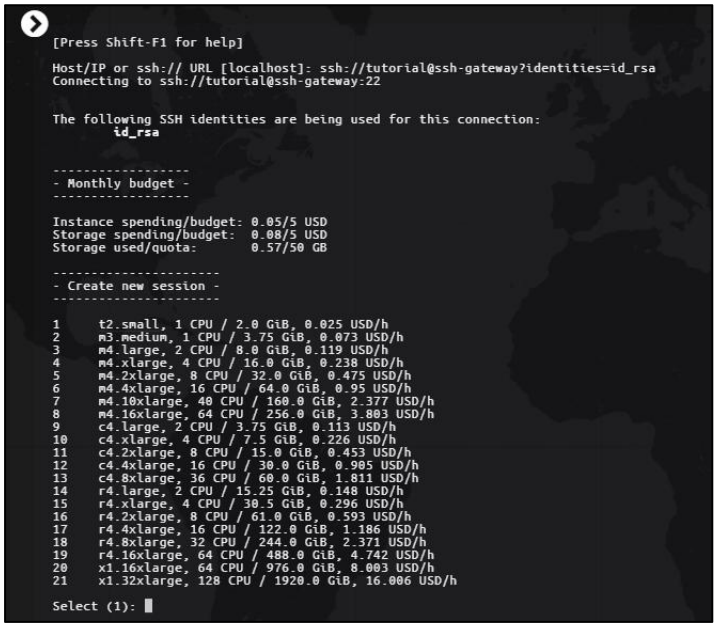
1. One of the ways SEPAL provides access to cloud computing infrastructure is through Amazon Web Services (AWS) cloud computing. In SEPAL you can select from a variety of cloud computers of various sizes. These cloud computers are referred to as **instances** and when they are switched on, you are running a **session**. An instance will automatically start if you launch an application in the processing options. The instance that automatically starts is always the smallest available one, the t2small, which has 1 GB of CPU and 2 GB of RAM power, for a small price of 0.03 USD per hour. If the session is not active for more than 15 minutes (i.e. you are not using any of the processing tools nor are you actively running anything in the terminal), the instance will be automatically shutdown.
2. You can manually shutdown an instance in your user resources, which you can access by clicking on your name.

3. Click the X next to the running instance to shutdown the existing instance.



B. Using the terminal to start an instance

1. First, go to the terminal 
2. If you don't have any instance running you should see the option to start a new session and the list of instances (see image below)



3. The time series analysis application is computationally intensive. You will need an instance with more CPUs and less memory because the algorithm is parallelized among many CPUs and does not require a huge amount of memory. Running a larger instance saves you time and money in your SEPAL account, depending on the size of the area. If the area is large, processing can take days to complete if using a small instance.

4. For running the test data set, you can first try selecting the c4.xlarge. If you are planning on running the application on your own area of interest, consider using a larger instance, such as c4.4xlarge or c4.8xlarge

5. To select the instance you need to type the number corresponding to the instance you would like to start. For example if you are starting the c4.xlarge, type **10** in the terminal and then hit the ENTER key on the keyboard. The command line only works by typing text in the terminal.

6. You will then see text about the amount of time you can leave your session running according to your user resources.

iv. If your user resources are limited you will be asked if you would like to proceed with starting this instance.

v. If you would like to proceed, type **y** (which stands for yes) in the terminal. Otherwise, if you would like to select another instance type **n** (which stands for no)**.**

vi. If you have plenty of resources you will not see this text.

7. It can take a minute to start an instance. You will see text that says 'Please wait….'

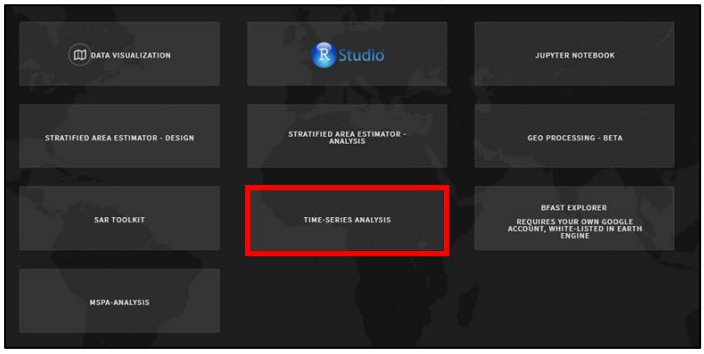
8. When the session has completed loading you will see your username@numbersandletters$

You have started an instance using the terminal! Now you can proceed you using the time series application at full power.

Part 3: Time series analysis app

A. Open the time series application

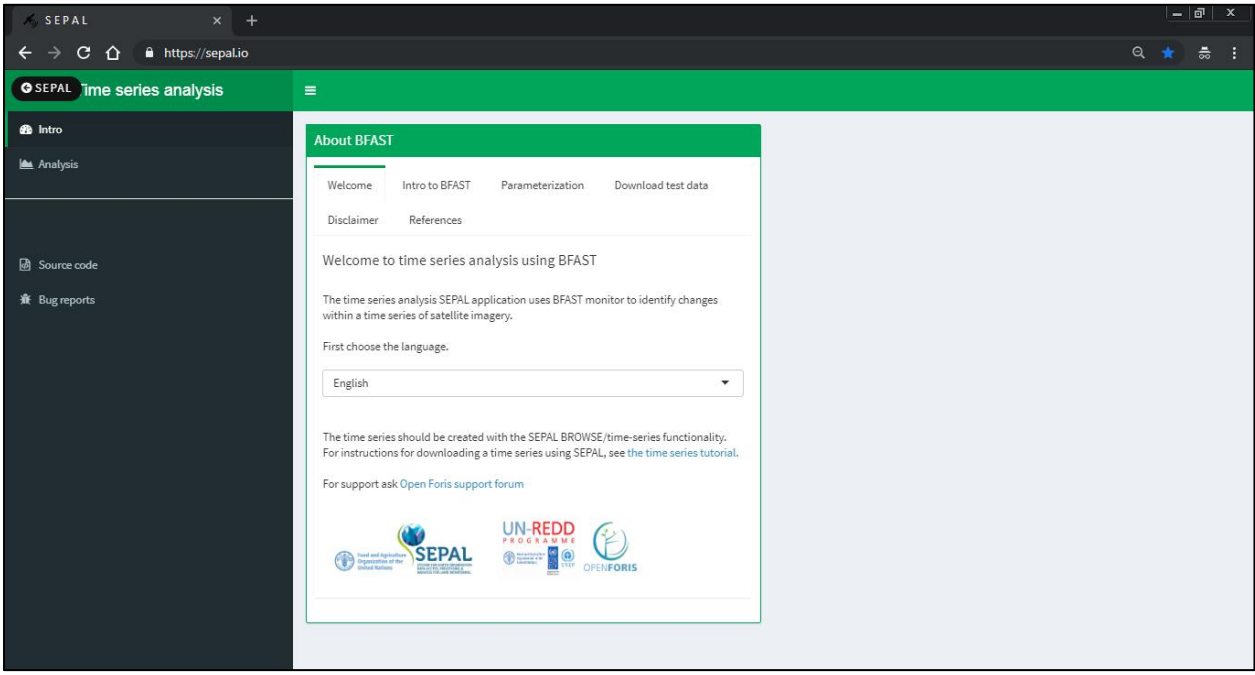
1. Go to the processing tools
2. Select **Time series analysis**



3. The application is split between 2 pages, the introduction and the analysis.

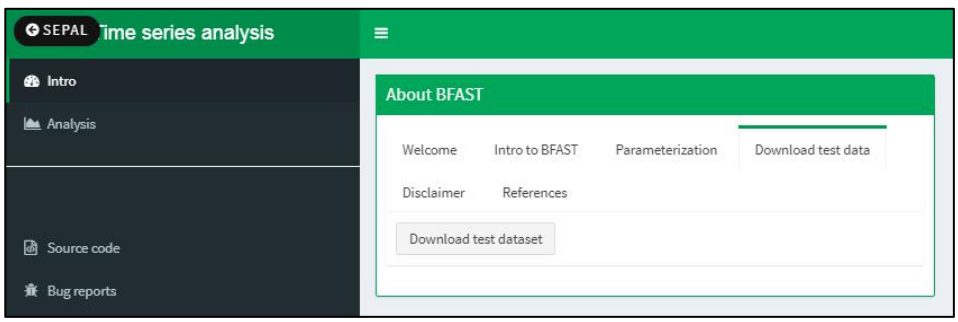
4. In the introduction section: the welcome, intro to BFAST and parameterization tabs provide additional information about the mechanics of the algorithm. It is recommended to read the text in these tabs.

Note: the application is only available in English at the moment. French and Spanish translations will be updated shortly. If you are interested in translating the application into your own language contact remi.dannunzio@fao.org or yelena.finegold@fao.org

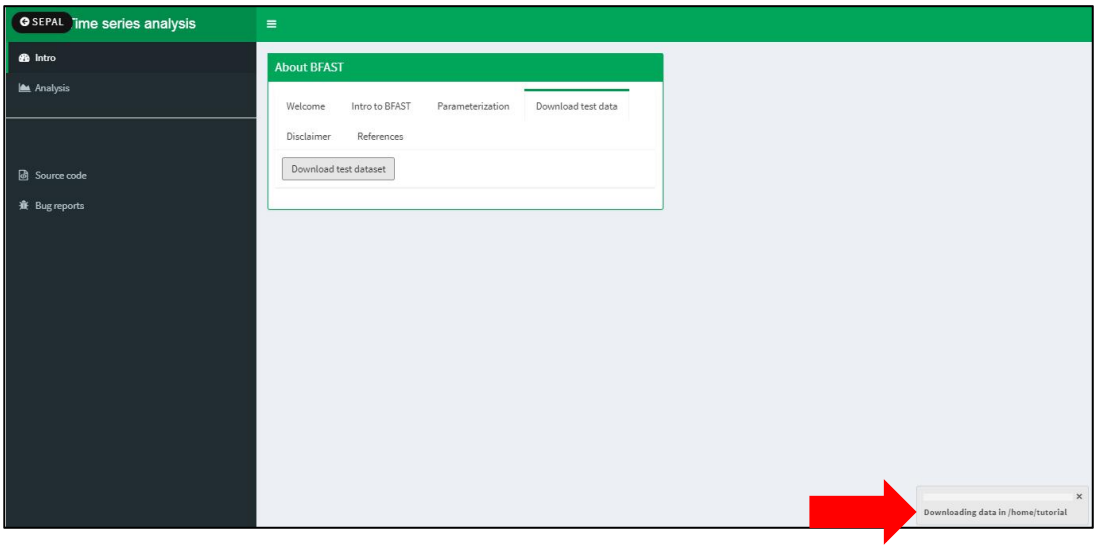


B. Download the test data set

1. If this is your first time using the time series analysis tool it is recommended that you first try with the example data set.The example data set can be downloaded in the '*download test data'* tab.

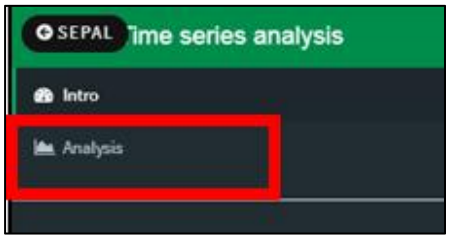


i. Click on the ‘Download test dataset’ button. The data is downloaded into the bfast_data_test folder in your root directory. The file location and information about the download will appear in the bottom right corner.

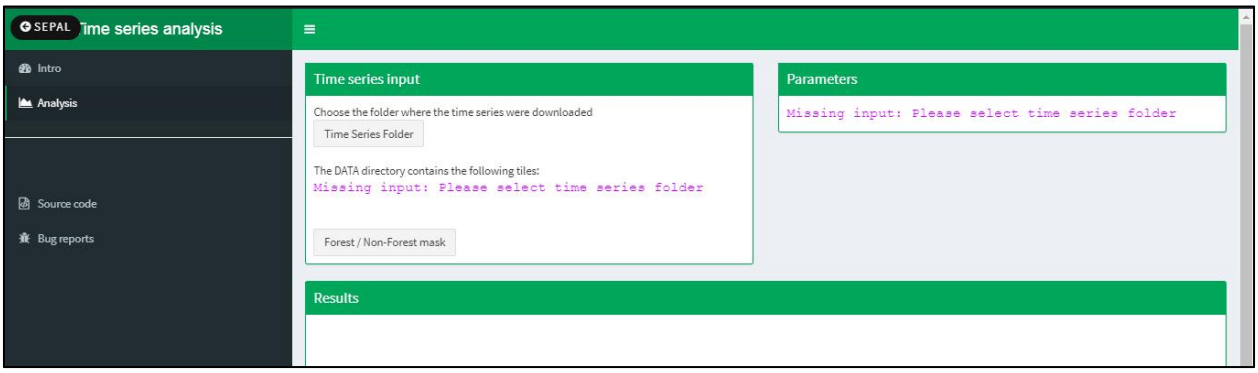


C. Run time series analysis over your area of interest

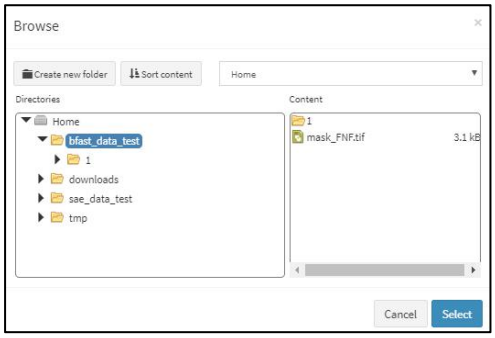
1. Click on the Analysis tab in the left column



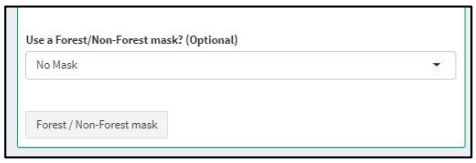
2. First you need to select the time series folder. Click on the Time Series Folder button to navigate to the folder with your downloaded data (either downloaded from the SEPAL search option or the test data set)



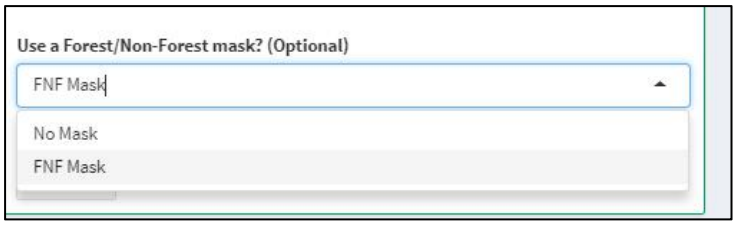
3. The file structure will always be timeseriesdownload/1/. Select the timeseriesdownload and not the numerical folder. In the case of the example dataset, select the bfast_data_test folder which is downloaded in your home drive.

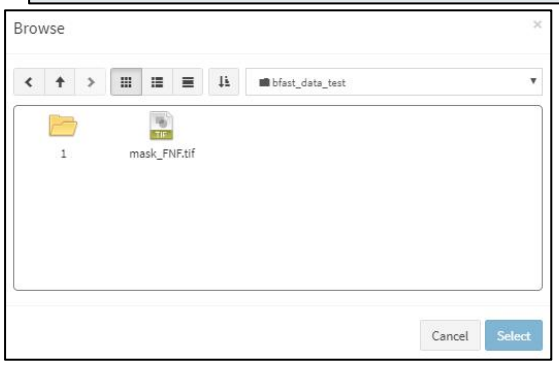


4. There is an option apply a mask and run BFAST only on areas outside the mask. You can select a file with 0 and 1 values. 0 values will be excluded and 1 included in the computation.



i. If you would like to use a mask, select **FNF mask** and then select the raster file by clicking on the **forest/non-forest mask** button and navigating to and selected the mask file.





5. Next change the parameters for your study area. If you are not sure which parameters to use follow Exercise 2, where you can run BFAST monitor for 1 pixel to test different parameters. The parameters include:

   i. **History beginning year**–The year that marks the start of the historical period. The actual start date will depend on the history parameter chosen.

   ii. **Monitoring start and end years**–The monitoring start year is the year that marks the end of the historical period and the start of the monitoring period. The monitoring end year marks the end of the monitoring period.

   iii. **History parameter**–Specifies the start of the stable history period. The options are:

a. reverse-ordered CUSUM (ROC), looks backward in time, using a stepwise approach, to identify a stable history period
b. Bai and Perron breakpoint estimation (BP), also identifies a stable history period and can additionally be used to identify disturbances in the history period.
c. all, uses all available observations.
d. numeric, i.e., 2011 , the start date can be specified using the year.

iv. **Elements of the formula**–The formula describes the type of regression model applied. The options are:
    a. trend + harmon, a linear trend and a harmonic season component
    b. harmon, a harmonic season component
    c. trend, a linear trend

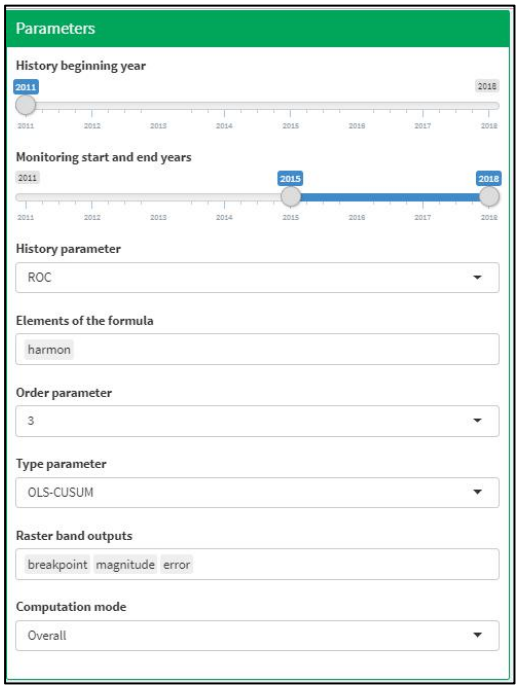
v. **Order parameter**–Specifies the order of the harmonic term, defaulting to 3.

vi. **Type parameter** –Specifies the type of monitoring process. For additional documentation on the type parameter see the strucchange package documentation (strucchange package documentation.) The options are:
    a. moving sums of residuals (MOSUM), where residuals are calculated as the difference between expected values and actual observations in a monitoring period based on OLS residuals.
    b. cumulative sum (CUSUM), cumulative sums of standardized residuals (MOSUM uses a moving sum, while CUSUM uses a cumulative of the same residuals)
    c. moving estimates (ME), the moving estimates process is returned
    d. fluctuation, returns the recursive estimates process

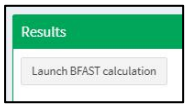
vii. **Raster band outputs**–Result layers to be returned. Can be any combination of *breakpoint, magnitude, error, history, r.squared, adj.r.squared, coefficients*. By default, *breakpoint, magnitude* and *error* are returned by the function. It is important to know which layers have been requested and in which order they will be exported because the layer names are not specified. Note that if "*coefficients*" is included, the output will include the following: "(*Intercept*)" and any trend and/or harmonic coefficients depending on the values of *formula* and *order*.

viii. **Computation mode**–chose between running the calculation for the entire monitoring period (overall) or for each year of the monitoring period (sequential)
    a. Overall-runs BFAST one time for the monitoring period and provides maximum one breakpoint for the entire monitoring period
    b. Sequential-runs BFAST for each year of the monitoring period. The output will be per year of the monitoring period

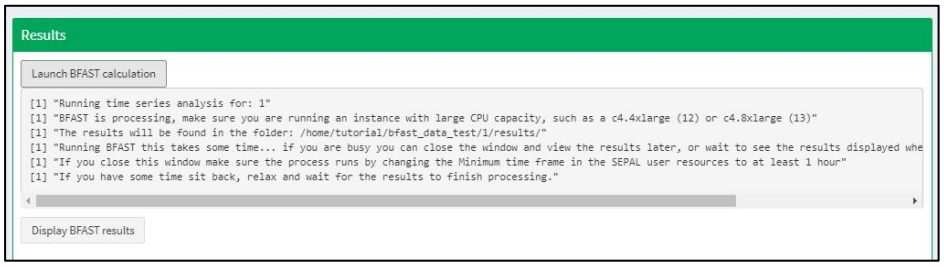
and provides maximum one breakpoint per year in the monitoring period. This option does not create the thresholded output and will not display the output within the application. To view the results use the visualizer in SEPAL or download the results to your local computer.



6. Once you have decided on your parameters, run BFAST by clicking on the **Launch BFAST calculation** button in the results box.

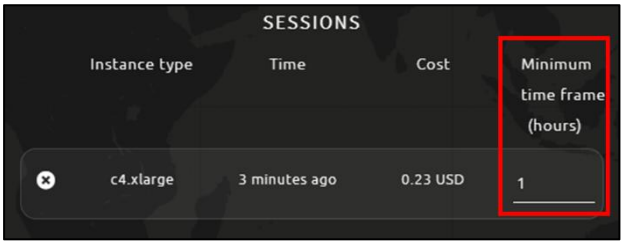


7. Depending on the size of your area and the size of your instance, BFAST can take a long time run. It is not necessary to keep this application open for the results to be created, it is only necessary to make sure that the instance is running.

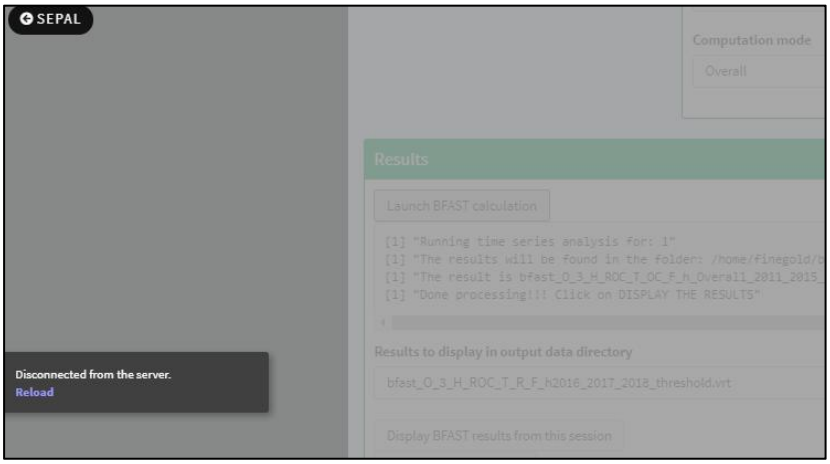


i. If your AOI has multiple polygons and contains many numeric folders, i.e. 1, 2, 3, etc, it will run the BFAST calculation for each of the folders, recursively.

ii. If you are running a large area or have a weak internet connection which might cause the application to disconnect you can go to your user resources in SEPAL and set the amount of time your session should stay open (see image below). This way you can shutdown SEPAL and the calculation will continue.

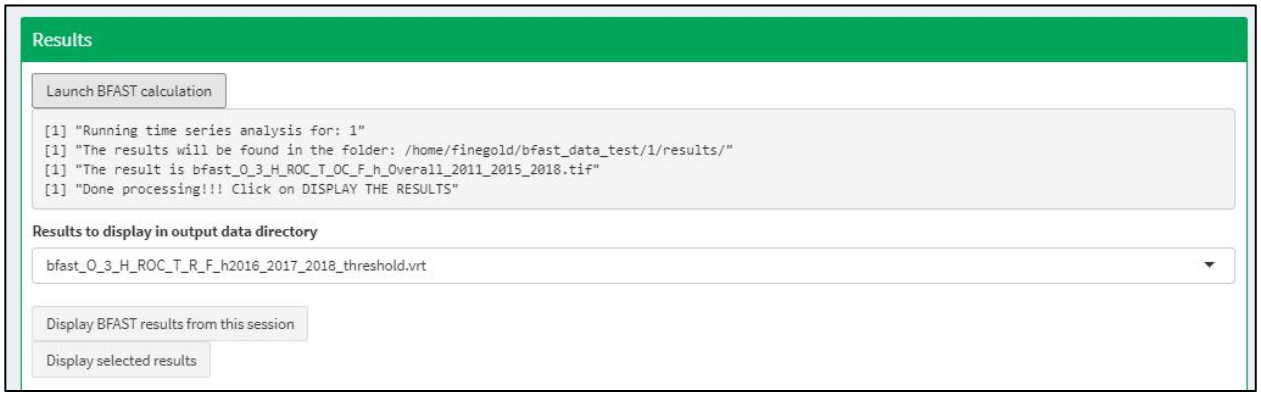


iii. If the page goes gray and you see **Disconnected from the server**, don't worry! The process is still running and you can follow the previous step to make sure your session remains active.

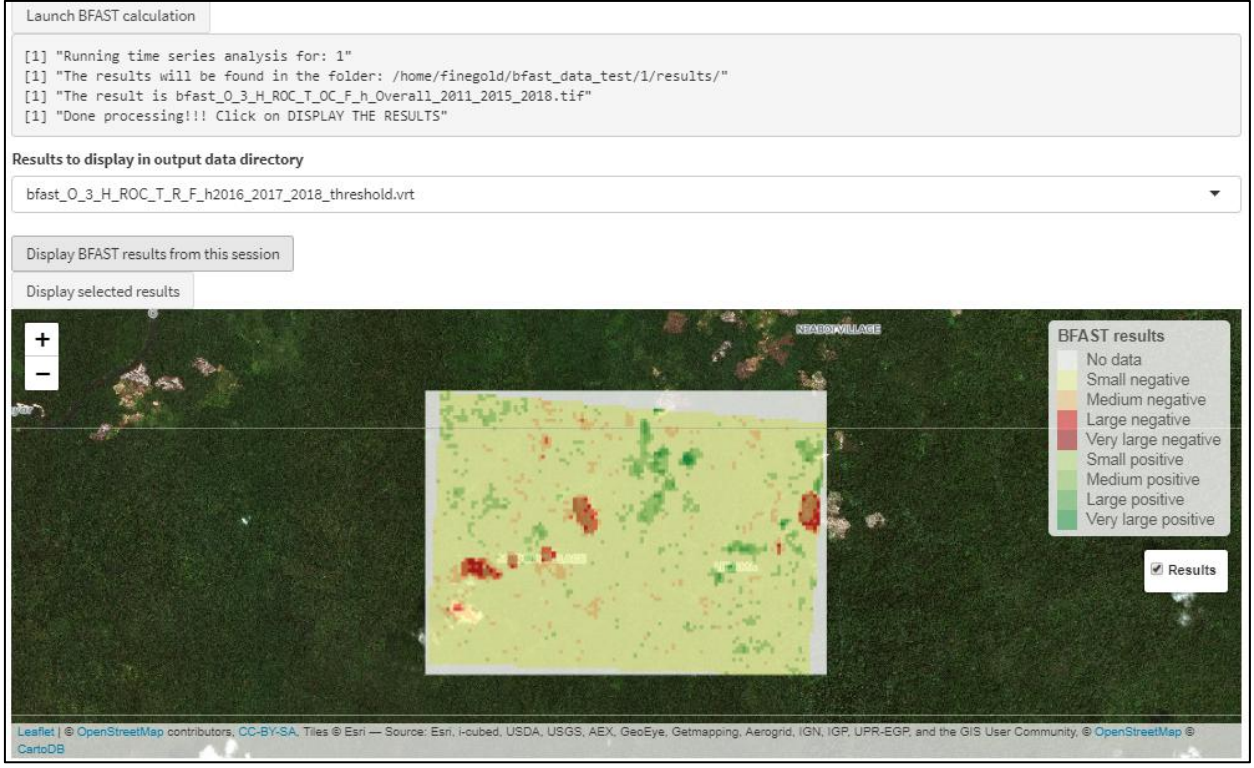


8. If you are feeling patient or have a small study area, you can wait for the algorithm to finish running and view one of the outputs, the thresholded magnitude.

9. When the calculation finishes running you will see the text **"Done processing!!! Click on DISPLAY THE RESULTS"**. You can now click on the **Display BFAST results from this session** button to display the thresholded magnitude.



10. The output from BFAST by default include 3 bands, the breakpoint, the magnitude and error (see 5vii). An additional output is calculated in this application, which is the thresholded magnitude. The thresholded magnitude is calculated using the magnitude output, calculating the mean magnitude value over the AOI and applying thresholds up to +/- 4 standard deviations from the mean.


Launch BFAST calculation
[1] "Running time series analysis for: 1"
[1] "The results will be found in the folder: /home/finegold/bfast_data_test/1/results/"
[1] "The result is bfast_O_3_H_ROC_T_OC_F_h_Overall_2011_2015_2018.tif"
[1] "Done processing!!! Click on DISPLAY THE RESULTS"
Results to display in output data directory
bfast_O_3_H_ROC_T_R_F_h2016_2017_2018_threshold.vrt
Display BFAST results from this session
Display selected results
BFAST results
No data
Small negative
Medium negative
Large negative
Very large negative
Small positive
Medium positive
Large positive
Very large positive
Results
Leaflet | © OpenStreetMap contributors, CC-BY-SA, Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community, © OpenStreetMap © CartoDB