D9 SERVICE SW DOCUMENTATION

EO AFRICA NATIONAL INCUBATORS (ANIN)

Prepared by:	VITO
Approved by:	Juan Suarez
Authorized by:	Juan Suarez

Code:	D9
Version:	v1
Date:	21/10/2024
Internal code:	GMV N/A

DOCUMENT STATUS SHEET

Version	Date	Pages	Changes
V0.1	20/09/2023	17	Population template
V0	03/11/2023	18	Review and finalization of V0.1
V1.0	21/08/2024	20	Final version of the document

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1. INTRODUCTION

1.1. PURPOSE

[Deliverable-9] Service SW documentation V0 is the first iteration of D9 delivered at month 14 of the ANIN project and builds on top of D8 Algorithm Theoretical Baseline Document [RD.1]. D9 summarizes all documentation associated to the developed of the drought monitoring pipelines integrated within openEO (e.g., specifications, architectural design, processing workflows, configuration files, libraries, software reused file, cloud-environment integration manuals).

1.2. SCOPE

D9 Service Software Documentation describes:

- Access to openEO
- The implementation of the workflows
- The dependencies: libraries, components and processors
- The input data needed to run the workflows

1.3. DEFINITIONS AND ACRONYMS

1.3.1. DEFINITIONS

Concepts and terms used in this document and needing a definition are included in the following table:

Table 1-1 Definitions

Concept / Term	Definition

1.3.2. ACRONYMS

Acronyms used in this document and needing a definition are included in the following table:

Table 1-2 Acronyms

Acronym	Definition
ANIN	South African Drought Monitoring
AOI	Area Of Interest
ASAP	Anomaly hot Spots of Agricultural Production
ATBD	Algorithm Theoretical Basis Document
CCI	Climate Change Initiative
CDI	Combined Drought Indicator
EOP	Earth Observation Programmes
EOS	End Of the growing Season
ESA	European Space Agency
FAO	The Food and Agriculture Organization
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
GDO	Global Drought Observatory
GLASS	The Global Land Surface Satellite
IDI	Integrated Drought Index
JRC	Joint Research Centre

MODIS	Moderate-Resolution Imaging Spectroradiometer
NDVI	Normalised Difference Vegetation Index
OLCI	SENTINEL-3 Ocean and Land Colour Instrument
PET	Potential Evapotranspiration
SGI	Standardised Groundwater Index
SMA	Soil Moisture Anomaly
SOS	Start Of the growing Season
SOW	Statement Of Work
SPEI	Standardised Precipitation Evapotranspiration Index
SPI	Standardised Precipitation Index
SSI	Standardised Streamflow Index
STD	Standard Deviation
VCI	Vegetation Condition Index

2. REFERENCES

2.1. APPLICABLE DOCUMENTS

The following documents, of the exact issue shown, form part of this document to the extent specified herein. Applicable documents are those referenced in the Contract or approved by the Approval Authority. They are referenced in this document in the form [AD.x]:

Table 2-1 Applicable Documents

Ref.	Title	Code	Version	Date
[AD.1]	Statement of Work. EO AFRICA - NATIONAL INCUBATORS EXPRO+	ESA-EOP-SD-SOW-0249	1.0	26/10/2021
[AD.2]	Proposal EO AFRICA NATIONAL INCUBATORS (ANIN)	GMV10277/22 V1/22,	1.0	18/02/2022

2.2. REFERENCE DOCUMENTS

The following documents, although not part of this document, amplify or clarify its contents. Reference documents are those not applicable and referenced within this document. They are referenced in this document in the form [RD.x]:

Table 2-2 Reference Documents

Ref.	Title	Code	Version	Date
[RD.1]	Algorithm Theoretical Baseline Document/s (ATBD) [Deliverable-8]	[Deliverable-8]	1.0	21/10/2024
[RD.2]				
[RD.3]				

3. DOCUMENT CONTENT

We structured the document into three main chapters dedicated to:

- the introduction of openEO and access to the NoR resources, and
- the implementation of drought monitoring pipelines, documentation of dependencies and input data.
- STAC data catalogue for ANIN.

4. OPENEO IMPLEMENTATION

The openEO Platform combines the capabilities of the depicted providers (**Error! Reference source not found.**) under a common interface: the openEO API. Which provides programming libraries to process a wide variety of earth observation datasets. This allows use cases from explorative research to large-scale production of EO-derived maps and information.

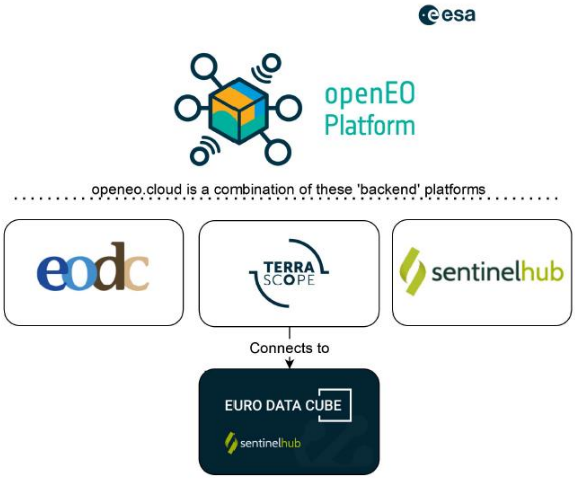


Figure 1. openEO providers

4.1. RESOURCES

Through the ESA Network of Resources (NoR) a sponsoring was granted which allowed for the provision of It resources on EODC – openEO platform. The NoR sponsoring included two one year subscriptions for advanced accounts, with the following specifications (Table 3).

Table 3. openEO account specifications

Credits/Month	CPUs	GB RAM	GB SSD	GB HDD
60,000	8	24	240	500

4.2. ACCESS

To gain access to openEO new ANIN users were requested to create a free trial account at https://docs.openeo.cloud/join/free_trial.html. This process is well documented and can be done through an EGI check-in. (The free trial is enough to run all layers.) Once these accounts were made, invitations to join EOPlaza[1] were sent out. EOPLaza is used to manage IT resources, provided under the NoR. Table 4 shows the list of active ANIN developers who have access to the openEO and the IT resources.

Table 4. Developers with access to openEO.

Name	Email	Organization
Emile Sonneveld	emile.sonneveld@vito.be	VITO
Jeroen Dries	jeroen.dries@vito.be	VITO
Enes Hisam	enes.hisam@gmv.com	GMV
Elia Cantoni I Gomez	elia.cantoni.igomez@gmv.com	GMV
Pablo Jesús Torres Hernández	pjtorres@gmv.com	GMV

[1] https://portal.terrascope.be/

4.2.1. DROUGHT MONITORING PIPELINES

The drought monitoring pipelines have been implemented into openEO. An overview of the ANIN drought monitoring setup with the individual implemented drought monitoring indices, their dependencies and input data are displayed at Figure 2. The dependency between SMA and SMI is not defined in openEO, but are strongly related. Note that SSFI, SGI and IDI are not mentioned in detail in this document. These layers do not profit from an openEO implementation as they are light enough to not scale well on a distributed system, and the first 2 do not involve satellite images.

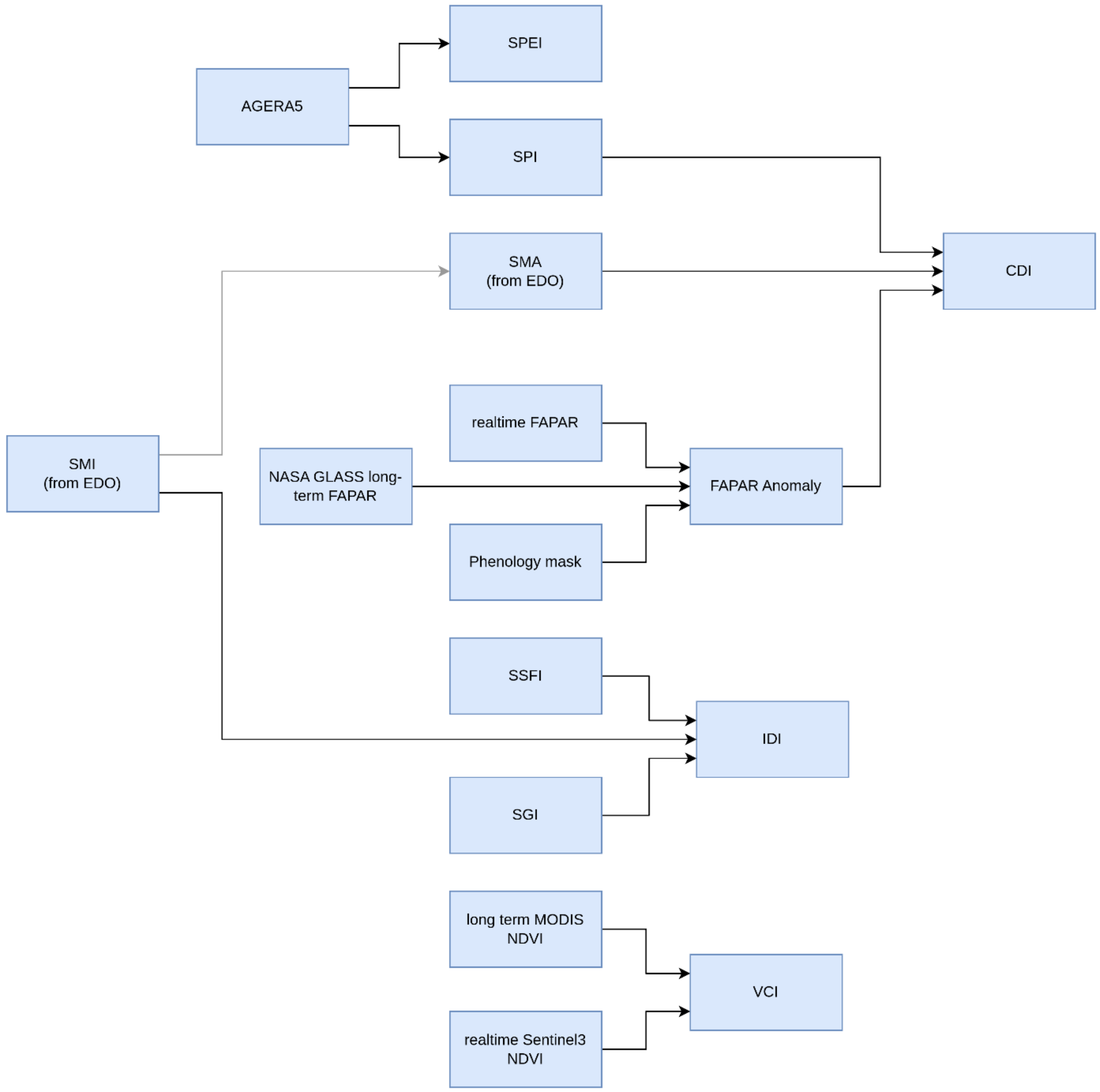


Figure 2. Flowchart of the ANIN drought monitoring system.

4.2.2. GIT REPOSITORY

The open source code is available under git here: https://github.com/VitoTAP/ANIN-drought-indices (The original GMV code can be found in a private repo here: https://git-ext.gmv.com/anin-external/drought-indices) The instructions on how to get the project running are described in the README. At the moment of writing, the main command to run is the industry standard: "**python3 -m pip install -r requirements.txt**".

In the git repository, the code for each layer can be found in files like 'LAYER/LAYER_openeo.py'. To get results of such a layer, one can run '***python3 –m LAYER/LAYER_openeo.py'***. The result will be outputted next to the file in a folder like '***out-2023-10-12_19_07_58.318432'.*** This makes it easy to run any layer, even the intermediate ones. By using the date in the output folder name, a layer can be calculated multiple times without mixing up the results. The outputs can consist of a couple of GeoTIFFs, accompanied by a job-results.json that contains STAC data for easy integration in different services. Example output:

Name	Size
job-results.json	541.9 kB
VCI_openeo_2022-01-01Z.tif	959.4 kB
VCI_openeo_2022-02-01Z.tif	915.9 kB
VCI_openeo_2022-03-01Z.tif	893.2 kB
VCI_openeo_2022-04-01Z.tif	950.1 kB
VCI_openeo_2022-05-01Z.tif	904.5 kB
VCI_openeo_2022-06-01Z.tif	735.2 kB
VCI_openeo_2022-07-01Z.tif	636.5 kB
VCI_openeo_2022-08-01Z.tif	700.1 kB
VCI_openeo_2022-09-01Z.tif	726.1 kB
VCI_openeo_2022-10-01Z.tif	753.8 kB
VCI_openeo_2022-11-01Z.tif	1.0 MB
VCI_openeo_2022-12-01Z.tif	1.0 MB
VCI_openeo_2023-01-01Z.tif	863.5 kB

Figure 3. GeoTIFFs outputs from openEO.

It is also possible to request CSV files, by specifying the following argument: "--out_format=CSV".

Below is an example of results generated as a NetCDF. It contains standard STAC data that describes the files next to it. Note that the returned files are compressed. This saves space and bandwidth but gives images of different size.

Name	Size
job-results.json	433.9 kB
openEO.nc	9.7 MB

Figure 4. NetCDF generated in openEO.

The SPI and SMA layers also have a notebook implementation. This is for more scientific people to explore and edit the code. They can be found here: SPI/SPI_openeo.ipynb and SMA/SMA_openeo.ipynb. There is a small video available that goes over what is available: https://youtu.be/CRsMDJEGKKM.

In some cases, data from "/data/users/Public/" is used. To manage this data, we recommend to create an account on Terascape and request a Virtual Machine (https://terrascope.be/en/form/vm). Through this online virtual machine, one can access this folder to add/remove/update any data that is present there. This is especially handy to import external python libraries, and additional collections that are not yet available by default in openEO.

4.3. USER DEFINED FUNCTIONS

For many layers, a User Defined Function (UDF) is used. The UDF is a way to call Python code in the openEO platform. The Python code accepts a small piece of the datacube, transforms it and returns a small datacube with the same shape. For example, we use an UDF in the *apply_dimension* process for SPI. Here we get as input a small surface area, across all dates. We pass this trough a python library and return the result.

A UDF allows for flexibility but comes with a performance hit, as Python code runs slower compared to doing the same logic with predefined openEO processes.

5. DROUGHT INDICATORS

5.1. METEOROLOGICAL DROUGHT

5.1.1. STANDARDISED PRECIPITATION INDEX (SPI)

Run with "**python3 -m SPI.SPI_openeo**"

The SPI layer is based on precipitation-flux layer from ERA5 Land. OpenEO already had this data from 2015 till the present time, but the GMV implementation uses data starting from 1960, so the ERA5 Land dataset[2] from Copernicus was imported and used via the Terrascope VM. In the imported dataset, no monthly aggregation is needed. The dataset is just loaded, and then, the SPI code from https://github.com/monocongo/climate_indices is called using Python in an UDF. Finally, the results are saved and downloaded.

The openEO process graph is shown in the next figure:

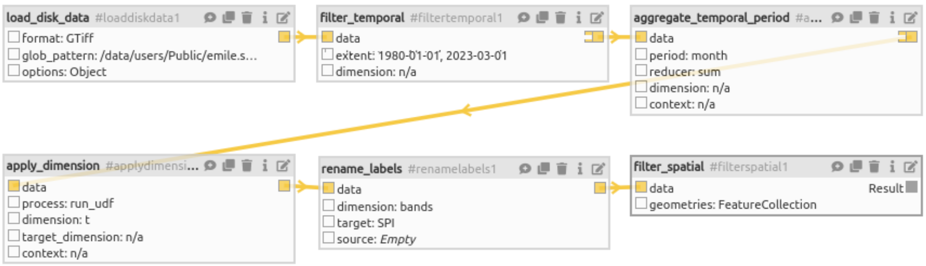
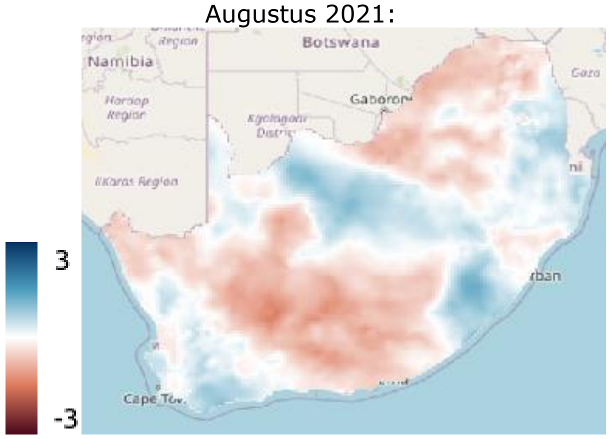
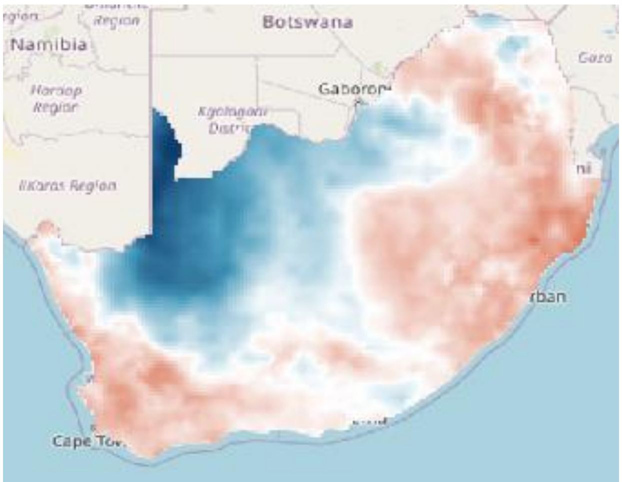


Figure 5. The SPI openEO process graph.

Table 5. Standardised Precipitation Index.

Product name	**STANDARDISED PRECIPITATION INDEX (SPI)**
Update frequency	Monthly with a delay of about three months relatively to actual date
Format	Raster format: NetCDF / GeoTIFF
Input data	Precipitation: ERA5-Land
Archive Length	January 1980 till 2024-05-01
Spatial resolution	0.1deg (native resolution: 9km)
Spatial coverage	Computed at national scale South Africa, Lesotho, and Eswatini
Requires Field Data	No

[2] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means

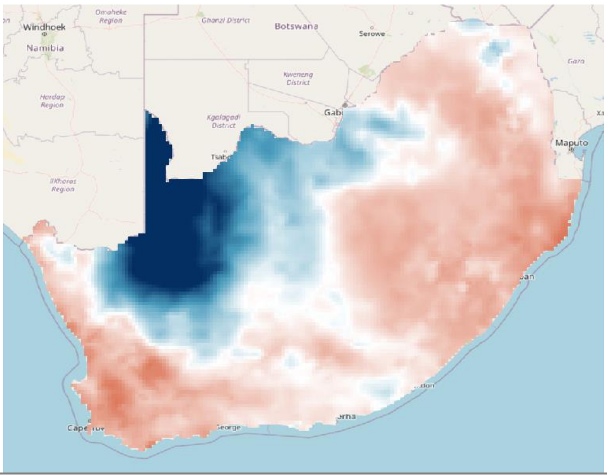
Product Overview	Augustus 2021:  Augustus 2022: 

5.1.2. STANDARDISED PRECIPITATION-EVAPOTRANSPIRATION INDEX (SPEI)

This is conceptually quite like the SPI. Also here, the newly imported ERA5 Land dataset is used. It is already provided on monthly basis, so there is no need to do an aggregate temporal. This time two bands are used: precipitation-flux and temperature-mean. They are loaded and passed directly to the UDF. Here, the SPEI code from https://github.com/monocongo/climate_indices is called to calculate the SPEI.

Table 6. Standardised Precipitation Evapotranspiration Index.

Product name	**STANDARDIZED PRECIPITATION-EVAPOTRANSPIRATION INDEX (SPEI)**
Update frequency	Monthly with a delay of about three months relatively to actual date
Format	Raster format: NetCDF / GeoTIFF
Input data	Precipitation: ERA5-Land
Archive Length	January 1980 to 2023-10-01
Spatial resolution	0.1deg (native resolution: 9km)
Spatial coverage	South Africa, Lesotho, and Eswatini
Requires Field Data	No
	Agustus 2022

5.2. SOIL MOISTURE DROUGHT (AGRICULTURAL/ECOLOGICAL)

5.2.1. VEGETATION CONDITION INDEX (VCI)

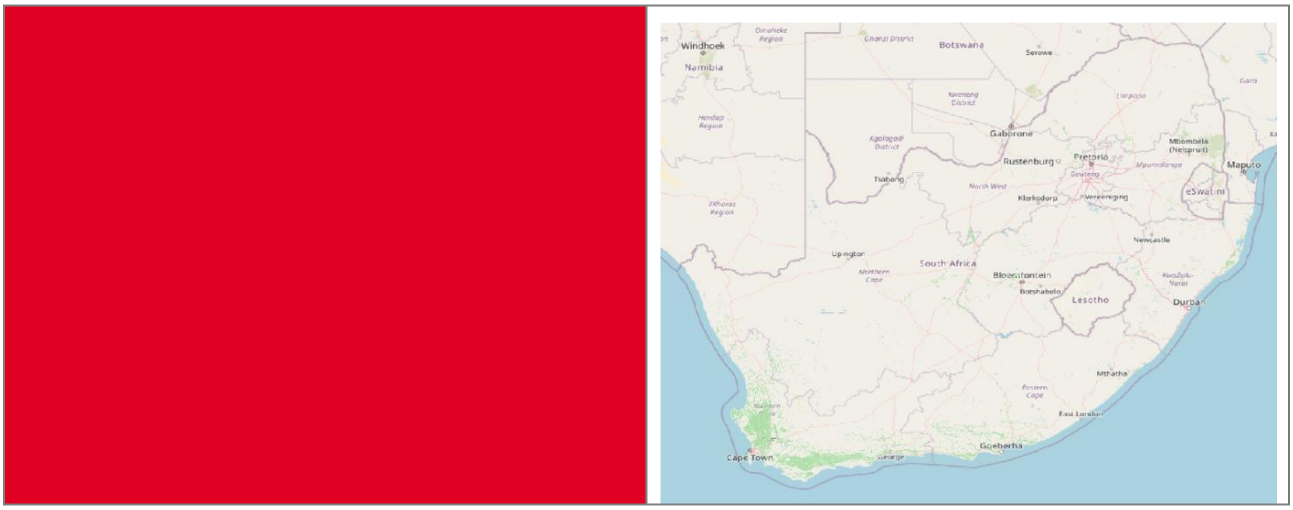
Run with "**python3 -m VCI.VCI_openeo**"

This layer compares current NDVI data to historical NDVI data on a month per month basis. Like other layers, basic pre-processing is done with standard openEO processes, and the rest is implemented in a UDF. The historical data is downloaded from the NASA's MODIS program (MOD13Q1) and the real-time data is retrieved from the already existing Copernicus Global Land Service NDVI layers in openEO. Both layers are first aggregated on monthly intervals and then passed through an UDF.

Table 7. Vegetation Condition Index.

Product name	**VEGETATION CONDITION INDEX (VCI)**
Update frequency	Monthly with a delay of 2 weeks
Format	Raster format: NetCDF / GeoTIFF
Input data	Satellite-based NDVI data: MODIS (MOD13Q1), Copernicus Global Land Service[3]S
Archive Length	July 2020 to Present
Spatial resolution	300m
Spatial coverage	South Africa, Lesotho, and Eswatini
Requires Field Data	No
	Augustus 2022

[3] https://land.copernicus.eu/global/products/ndvi



5.2.2. COMBINED DROUGHT INDICATOR (CDI)

Run with "**python3 -m CDI.CDI_openeo**"

The CDI layer is based on three in-between layers: SPI, SMA, FAPAR_Anomaly and the SPI of the previous month. SPI is already described in chapter 5.1.1. The in-between layers are described fist here followed by the CDI layer itself.

SMA:

Run with "**python3 -m SMA.SMA_openeo**"

The SMA layer was downloaded from JRC (https://edo.jrc.ec.europa.eu/gdo/php/index.php?id=2112) and imported in OpenEO. No clipping was done as the data is light enough, and OpenEO clips automatically when required.

SPI of previous month:

The SPI of the previous month is calculated by taking offsetting the SPI with −1 over the time dimension.

FAPAR anomaly:

Run with "**python3 -m FAPAR_Anomaly.FAPAR_Anomaly_openeo**"

For this layer, data from NASA GLASS was downloaded for long term statistics (http://www.glass.umd.edu/FAPAR/MODIS/250m/). For the near-real-time statistics, a CGLS layer that was already available in openEO was used: CGLS_FAPAR300_V1_GLOBAL.

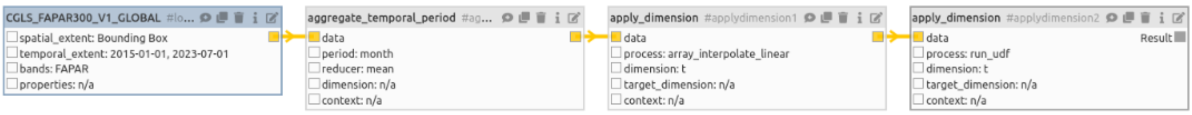


Figure 6. FAPAR anomaly process graph.

After loading the data, the days are aggregated to monthly intervals, using the mean to aggregate all data per month. Then the array_interpolate_linear function is called to fill in any missing data. Finally, the data is passed in an UDF and the result is saved.

This layer is masked with a per-month Phenology based layer. The 12 masks that were generated by GMV are used for this.

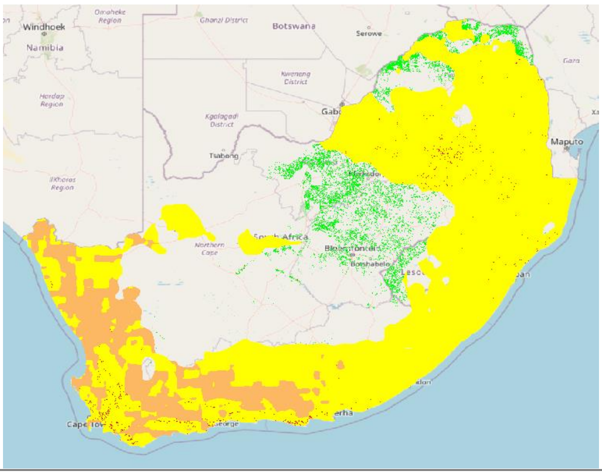
CDI

CDI is calculated based on thresholds of SPI, SMA, and FAPAR anomaly.

CDI classifies each pixel based on the 3 input layers as follows:

Level	CLASSIFICATION CONDITION
Watch	SPI-3 < -1
Warning	SMA < -1 and SPI-3 < -1
Alert	FAPAR anomaly < -1 and SPI-3 < -1
Partial recovery	FAPAR anomaly < -1 and SPI-3(m-1) < -1 and SPI-3 > -1
Full recovery	FAPAR anomaly > -1 and SPI-3(m-1) < -1 and SPI-3 > -1

Note that even though most of the FAPAR Anomaly layer is masked away, this does not mask away most of the CDI layer. As the data of this layer is not read for every level, and 'nan<1' gives 'False' instead of making outcome 'nan'.

Product name	**COMBINED DROUGHT INDICATOR (CDI)**
Update frequency	Monthly (with some delay)
Format	Raster format: NetCDF / GeoTIFF
Input data	• SPI • SMA • FAPAR Anomaly
Archive Length	2001 to 2024-04-01
Spatial resolution	300m
Spatial coverage	South Africa, Lesotho, and Eswatini
Requires Field Data	No
	November 2022 

6. STAC DATA CATALOG FOR ANIN

The output of openEO can consist of a couple of GeoTIFFs or a NetCDF, accompanied by a **job-results.json** that contains STAC data for easy integration in different services. Example output:

Name	Size
job-results.json	541.9 kB
VCI_openeo_2022-01-01Z.tif	959.4 kB
VCI_openeo_2022-02-01Z.tif	915.9 kB
VCI_openeo_2022-03-01Z.tif	893.2 kB
VCI_openeo_2022-04-01Z.tif	950.1 kB
VCI_openeo_2022-05-01Z.tif	904.5 kB
VCI_openeo_2022-06-01Z.tif	735.2 kB
VCI_openeo_2022-07-01Z.tif	636.5 kB
VCI_openeo_2022-08-01Z.tif	700.1 kB
VCI_openeo_2022-09-01Z.tif	726.1 kB
VCI_openeo_2022-10-01Z.tif	753.8 kB
VCI_openeo_2022-11-01Z.tif	1.0 MB
VCI_openeo_2022-12-01Z.tif	1.0 MB
VCI_openeo_2023-01-01Z.tif	863.5 kB

Below is an example of results generated as a NetCDF:

Name	Size
job-results.json	433.9 kB
openEO.nc	9.7 MB

It is a STAC collection that contains items that describes the files next to it. It also keeps track what products were used to generate this output, providing out-of-the-box attribution to data providers. An item out of this collection may look like this:


```
"VCI_openeo_2012-01-01Z.tif": {
  "eo:bands": [
    {
      "name": "VCI"
    }
  ],
  "file:nodata": [
    255
  ],
  "file:size": 56908,
  "href": "https://openeo-dev.vito.be/openeo/1.1/jobs/j-64543cc560934b93b8752784f193df10/results/assets/YzkwNzQyODYzMmI0NmNiNDA2ZWYwMmEwMWE0YTVkYzM0YWVkYWQzNzAzNjlkN2I5ZDVjMzE0YTNiNjY2ZmQwM0BlZ2kuZXU%3D/f8ca614ba5c8604b983ac00cccae2057/VCI_openeo_2012-01-01Z.tif?expires=1698186856",
  "proj:bbox": [
    26.9985119,
    -27.0014881,
    30.0014881,
    -25.9985119
  ],
  "proj:epsg": 4326,
  "proj:shape": [
    1009,
    337
  ],
  "raster:bands": [
    {
```



```
        "name": "NDVI",
        "statistics": {
          "maximum": 214.0,
          "mean": 144.68284460219,
          "minimum": 1.0,
          "stddev": 21.808507837379,
          "valid_percent": 99.1
        }
      }
    ],
    "roles": [
      "data"
    ],
    "title": "VCI_openeo_2012-01-01Z.tif",
    "type": "image/tiff; application=geotiff"
}
```


One can visualise the STAC catalogue with external tools like radiantearth's stac-browser. The following URL is an example of how to visualise an openEO output with this tool:

https://radiantearth.github.io/stac-browser/#/external/openeo.vito.be/openeo/1.1/jobs/j-231030e45df84bbdb6a66148f3f493a1/results/YzkwNzQyODYzMmI0NmNiNDA2ZWYwMmEwMWE0YTVkYzM0YWVkYWQzNzAzNjlkN2I5ZDVjMzE0YTNiNjY2ZmQwM0BlZ2kuZXU%3D/e8ff3d789e7564a173484c345b96a92e?expires=1699280564

To visualise the output, take the "canonical" url out of the "job_results.json" and use it to replace the last part of the stac-browser url.

END OF DOCUMENT