

From raw data to biodiversity indicators



DATA
TERRA

Marie Jossé



Pôle National
de Données de Biodiversité

AGENDA

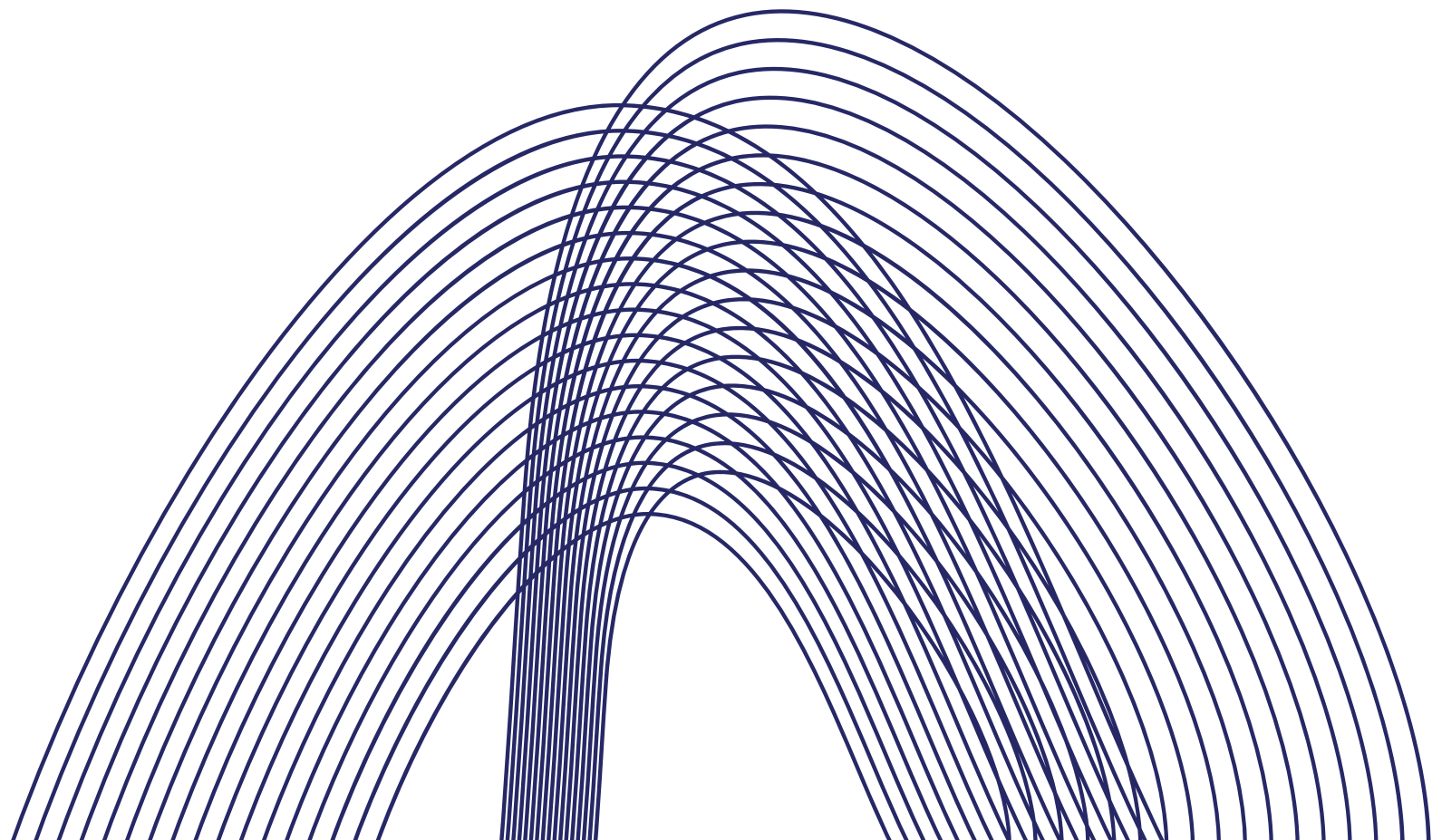
What are we talking about

01

Remote sensing
with Sentinel 2

02

Biodiversity data
exploration an
overview

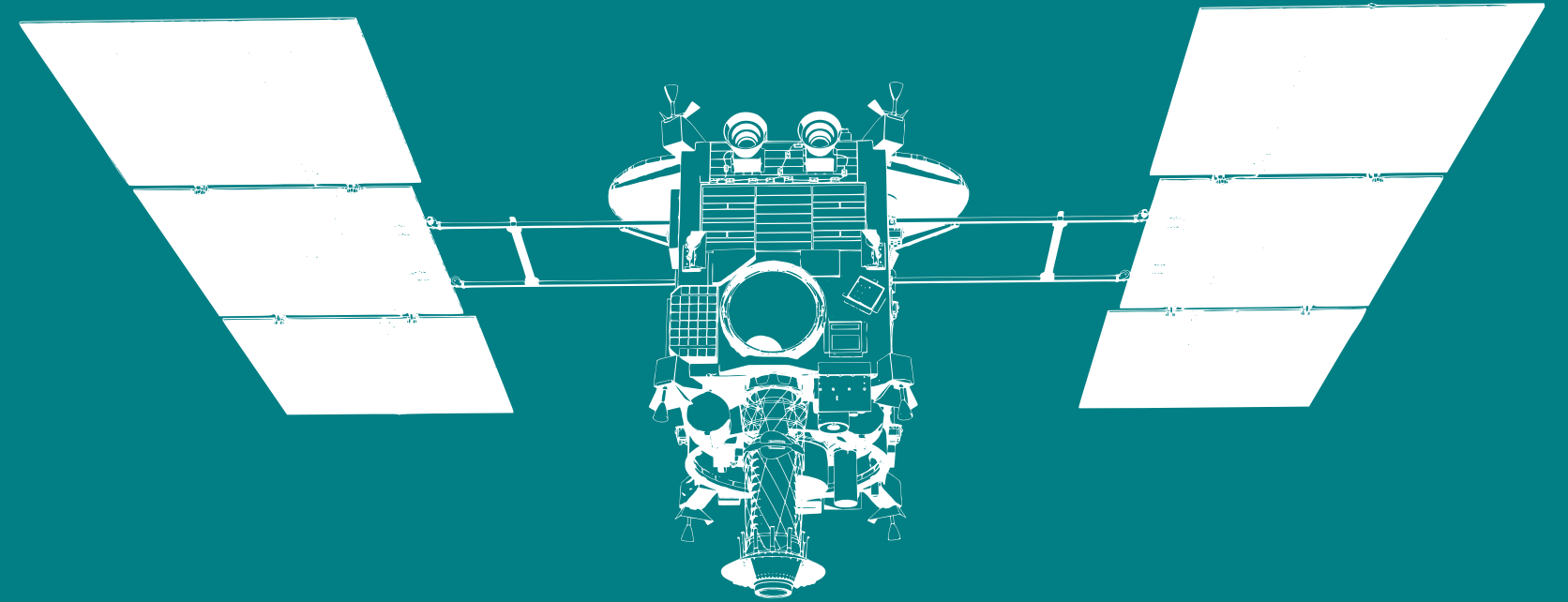


Context

Remote sensing

For biodiversity indicators

What's the point ?



Download S2 data

Multiple platforms :
SciHub, Peps and Theia



A screenshot of the Copernicus Open Access Hub search interface. The browser address bar shows 'scihub.copernicus.eu/dhus/#/home'. The page header includes the ESA and Copernicus logos and the text 'Copernicus Open Access Hub'. The main content area features a search bar and a filter panel on the left. The filter panel is set to 'Mission: Sentinel-2' and shows the following options: Satellite Platform: 'S2A_*', Product Type: 'S2MSI2A', Relative Orbit Number (from 1 to 143): empty, Cloud Cover % (e.g. [0 TO 9.4]): empty. The main area displays a map of Europe and surrounding regions, with various cities and countries labeled. The map is overlaid with a grid of satellite swaths.

Preparing your data

Preprocess Sentinel 2

Upload your data

SciHub a zip folder

Prepare

The tool reads and give 2
files to use for your
analysis

Raster Bil

Image stored as a
succession of pixels

Metadata

Description of the raster
bil

Envi Bil Format

Quick overview of what the tool looks like

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy Configured by Planemo' and various utility links. The left sidebar contains a 'Tools' section with a search bar and a list of tool categories, including 'Preprocess sentinel 2 data'. The main content area displays the tool's configuration page, which includes an 'Input data' field with a file list, a dropdown menu for 'Where does your data come from?' set to 'From Scihub or Peps', and an 'Execute' button. Below the configuration, the tool's documentation is visible, including a title, a description of what it does, an input description, and an output description.

Galaxy Configured by Planemo Workflow Visualize Bibliothèque de données Admin Aide Utilisateur

Tools search tools Upload Data

Preprocess sentinel 2 data read, crop, resample and write it as a raster stack. (Galaxy Version 0.0.0)

Input data

- 3: S2A_MSIL2A_20200306T015621_N0214_R117_T51JXN_20200306T034744.zip
- 2: S2A_Subset_test
- 1: S2A_Subset_test.hdr

Where does your data come from ?
From Scihub or Peps

Execute

Preprocess Sentinel 2 data

What it does

The goal of this tool (using the package `preprocS2`) is to provide a common framework for the preprocessing of Level-2A Sentinel-2 images (does not work yet for Level 1C). Sentinel-2 L2A images can be produced or obtained from various data hubs or atmospheric correction methods. `PreprocS2` provides a unique function to read, crop, resample the original image directory, and write it as a raster stack.

Input description

A zip folder with Sentinel 2 data. These data can be downloaded on 3 different platforms:

- Copernicus Open Access Hub, SciHub, which provides complete, free and open access to Sentinel-2 data : <https://scihub.copernicus.eu/dhus/#/home>
- PEPS, la 'Plateforme d'Exploitation de Produits Sentinel' : <https://peps.cnes.fr/rocket/#/search>
- Theia : <https://theia.cnes.fr/atdistrib/rocket/#/search>

For each of those 3 platforms you will need to create an account.

You need to select from where you downloaded your data in order for the tool to know the format of the folder once it unzips it.

ZIP	Source
folder.zip	Character
...	...

Output

A zip folder containing the reflectance data with a raster stack pile in BIL format (no extension), a header with the metadata (.hdr extension) and 2 other files in .xml.

A zip folder containing the cloud mask with the 2 files for the raw data (.RAW) and 2 files containing the metadata (.hdr).

Rasterdiv

Compute biodiversity indicators

Global diversity overview to find the hot spot

S2A_Subset	longitude	latitude	Shannon	Renyi	Berger-Parker	Pielou	Hill	Prao	CRE
231	13.7119569929834	3.17838623784837	2.60890630597545	2.60890630597545	0.16	356880	13.5841857815757	5.1584	356880
231	13.7137566357487	3.17838849139271	2.69060859631006	2.69060859631006	0.133333333333333	357080	14.7406442928311	5.10888888888889	357080
228	13.7155562797955	3.17839074178832	2.67674384766503	2.67674384766503	0.142857142857143	357280	14.5376792458669	4.84571428571429	357280
231	13.7173559251221	3.1783929890352	2.67118246534123	2.67118246534123	0.125	357480	14.4570540550398	4.785	357480
219	13.7191555717267	3.17839523313333	2.7307675777817	2.7307675777817	0.111111111111111	357680	15.3446607213507	5.04691358024691	357680
223	13.7209552196075	3.1783974740827	2.78790610193132	2.78790610193132	0.111111111111111	357880	16.2469646722969	5.26024691358025	357880
230	13.7227548687627	3.17839971188331	2.86114692117586	2.86114692117586	0.111111111111111	358080	17.4815654206707	5.32543209876543	358080
233	13.7245545191906	3.17840194653513	2.87277465770396	2.87277465770396	0.111111111111111	358280	17.6860228423754	5.67703703703704	358280
240	13.7263541708893	3.17840417803817	2.86114692117586	2.86114692117586	0.111111111111111	358480	17.4815654206707	5.71654320987654	358480
229	13.7281538238571	3.1784064063924	2.85359585287384	2.85359585287384	0.111111111111111	358680	17.350058061542	5.83703703703704	358680
227	13.7299534780921	3.17840863159782	2.77434222019488	2.77434222019488	0.088888888888889	358880	16.0280805797061	5.85679012345679	358880
224	13.7317531335926	3.17841085365442	2.77434222019488	2.77434222019488	0.088888888888889	359080	16.0280805797061	5.92592592592593	359080
230	13.7335527903567	3.17841307256218	2.92837492698598	2.92837492698598	0.088888888888889	359280	18.6972214448931	7.26716049382716	359280
234	13.7353524483827	3.17841528832109	2.98844975483482	2.98844975483482	0.111111111111111	359480	19.8548786913162	8.05333333333333	359480
236	13.7371521076688	3.17841750093115	2.99600082313684	2.99600082313684	0.111111111111111	359680	20.0053717129103	8.46617283950617	359680
229	13.7389517682132	3.17841971039234	3.02079455106064	3.02079455106064	0.088888888888889	359880	20.5075795316409	9.17728395061728	359880
237	13.7407514300141	3.17842191670466	2.9692709500047	2.9692709500047	0.111111111111111	360080	19.4777141909957	8.14222222222222	360080
235	13.7425510930697	3.17842411986808	2.91520893559028	2.91520893559028	0.111111111111111	360280	18.4526674189673	7.52987654320988	360280



Input

The 2 envi bil files from the first tool

Output

Multiple biodiversity indicators shown in tabular and in graphs

How does it look like ?

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy Configured by Planemo' and various utility links. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button. Below this, several tool options are listed, including 'Compute biodiversity indices from remote sensing data'. The main panel displays the configuration for this tool, with fields for 'In which format are your data?', 'Input raster' (set to '2: S2A_Subset_test'), 'Input header' (set to '1: S2A_Subset_test.hdr'), and a text input for 'alpha' (set to '1'). An 'Execute' button is visible below the configuration.

Compute biodiversity indices from remote sensing data (Galaxy Version 0.0.0)

In which format are your data ?
Your already have the files in ENVI BIL format

Input raster
2: S2A_Subset_test
It can be the raw data in bil or the PCA raster layer in bil

Input header
1: S2A_Subset_test.hdr

Write a number of the value of alpha you want to calculate Renyi, Prao and Hill indicators (between 0 and 5)
1

Execute

Computes global biodiversity indices from satellite remote sensing data

What it does
This tool aims to provide functions to apply Information Theory based diversity indexes on RasterLayer such as Shannon's entropy or Cumulative Residual Entropy (CRE).

Input description
It expects an image file as input, with a specific data format. ENVI HDR image with BIL interleave required. The image is an ENVI raster including :

- A binary file (which has no extension here).
- A header file (with .hdr extension).

The header file is a text file including all necessary metadata which can be read with a text editor. It includes image dimensions, projection, and the name and central wavelength for each spectral band.

In order to get such input we advise to use the tool preprocessing sentinel 2 data. If you did so you can directly enter the "Reflectance" output from this tool and thus select the option "The data you are using are in a zip folder Reflectance".

⚠ If you do not use this Reflectance folder make sure that your data are respectively in bil and hdr format in the datatypes.

- A number for the alpha indice which used to calculate the following indicators : Renyi, Hill and Prao.

BIL	ENVI HDR	Number alpha
raster stack	Metadata	1

BiodivMapR

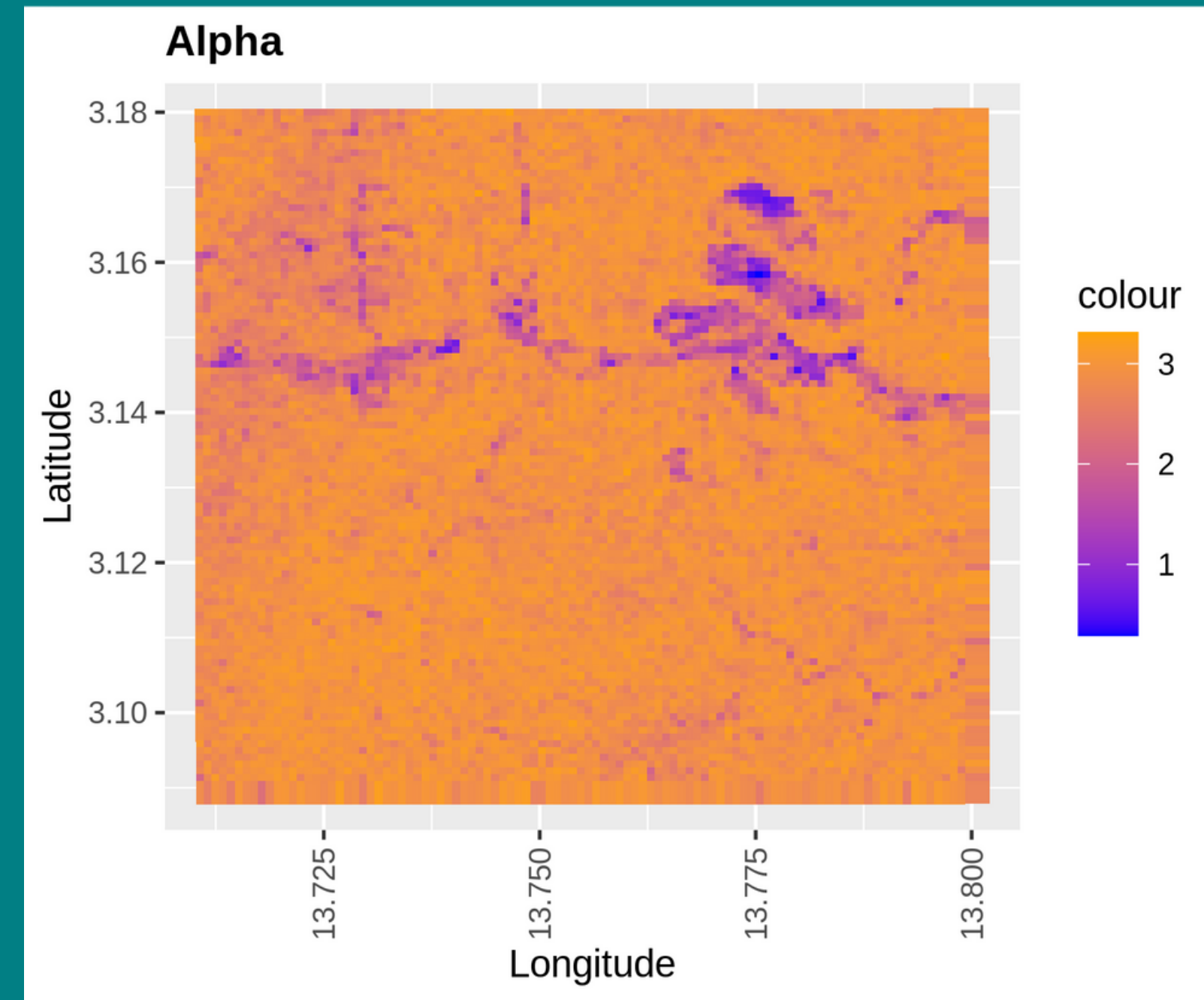
Canopy study :
Map diversity tool

Input

- The 2 envi bil files from the previous tool

Output

- Alpha diversity
- Beta diversity
- Functionnal analysis



Tools	Functionnal	longitude	latitude
search tools	0.0203434638679028	13.7111010965961	3.17924441566019
	0.0203434638679028	13.7111910787478	3.17924452851753
Upload Data	0.0203434638679028	13.7112810609028	3.179244641367
	0.0203434638679028	13.7113710430609	3.1792447542086

How does it look like ?

Map diversity

Galaxy Configured by Planemo | Workflow | Visualize | Bibliothèque de données | Admin | Aide | Utilisateur

Tools

search tools

Upload Data

Upload File from your computer

Compute biodiversity indices from remote sensing data

Compare diversity indicators with remote sensing data

Preprocess sentinel 2 data read, crop, resample and write it as a raster stack.

Compute a PCA raster from remote sensing data

Compute spectral indices as NDVI from remote sensing data

Validate ISO 19139 metadata documents from files or Urls.

Map diversity from remote sensing data

WORKFLOWS

All workflows

Map diversity from remote sensing data (Galaxy Version 0.0.0)

In which format are your data ?

Your already have the files in ENVI BIL format

Input raster

2: S2A_Subset_test

It can be the raw data in bil or the PCA raster layer in bil

Input header

1: S2A_Subset_test.hdr

Do you want to do a PCA or a SPCA ?

SPCA
 PCA

If you choose PCA there is no rescaling of the data as opposed as if you choose SPCA

Alpha, beta, functional diversity and comparison plot and map

Alpha diversity map
 Beta diversity map
 Functional diversity map
 All of the above

Execute

Process satellite remote sensing data to produce biodiversity indicators

What it does

Féret and Asner (2014) developed a method for **tropical forest** diversity mapping based on very high spatial resolution airborne imaging spectroscopy.

The goal of this tool using the package biodivMapR is to produce (spectral) diversity maps based on (optical) images.

Input description

It expects an image file as input, with a specific data format. ENVI HDR image with BIL interleave required. The image is an ENVI raster including :

- A binary file (which has no extension here).
- A header file (with .hdr extension).

The header file is a text file including all necessary metadata which can be read with a text editor. It includes image dimensions, projection, and the name and central wavelength for each spectral band.

BiodivMapR

Canopy study : Comparison tool

Input

- The 2 envi bil files from the first tool
- if you want to compare with in situ data you'll also need a shape file

ID_Plot	Forest_HighDiversity	Forest_LowDiversity	Forest_MediumDiversity	LowVegetation	Forest_HighDiversity	Forest_LowDiversity
Forest_HighDiversity	0	0.592972288569817	0.565601439639129	0.641352656343501	0.553116089728383	0.723287828668831
Forest_LowDiversity	0.592972288569817	0	0.794954586282086	0.781355810742138	0.68534833660536	0.782484730632053
Forest_MediumDiversity	0.565601439639129	0.794954586282086	0	0.576192328969609	0.497822379593039	0.649394360809528
LowVegetation	0.641352656343501	0.781355810742138	0.576192328969609	0	0.523405001569518	0.557895945457021
Forest_HighDiversity	0.553116089728383	0.68534833660536	0.497822379593039	0.523405001569518	0	0.569242046482229
Forest_LowDiversity	0.723287828668831	0.782484730632053	0.649394360809528	0.557895945457021	0.569242046482229	0
Forest_MediumDiversity	0.994515267618716	1	0.998625429553265	1	0.996818499650385	0.998657718120805
LowVegetation	0.94018897315834	0.95656976529877	0.989232155773878	0.994662436357695	0.968597398220022	0.958651854571807
Forest_HighDiversity	0.955115603705467	0.99864406779661	0.968903005991497	0.988401708610167	0.964193053605717	0.97430124627628
Forest_LowDiversity	0.951684436546145	0.998625429553265	0.961175350935895	0.986886991391772	0.959473727734387	0.975508315582341

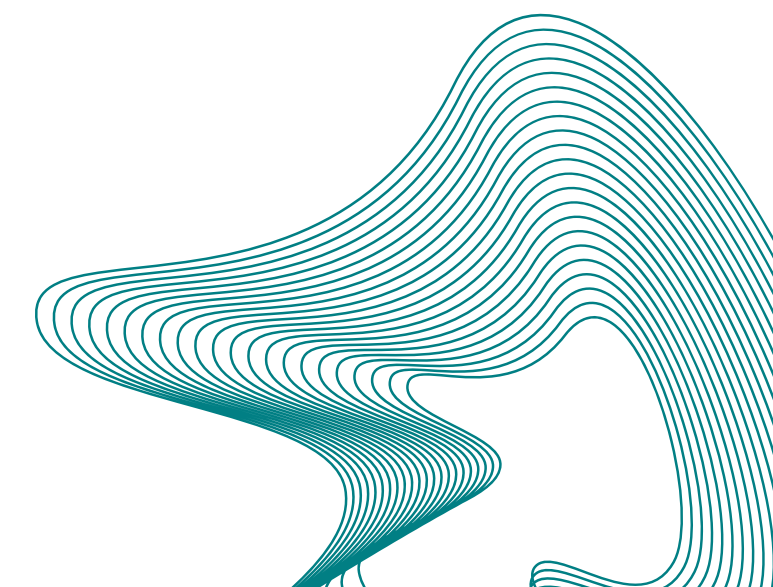
Output

- Bray curtis table
- comparison study

How does it look like ?

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy Configured by Planemo' and various menu items like 'Workflow', 'Visualize', and 'Admin'. The left sidebar contains a 'Tools' section with a search bar and an 'Upload Data' button, followed by a list of tool categories such as 'Upload File from your computer', 'Compute biodiversity indices from remote sensing data', and 'Map diversity from remote sensing data'. The main content area displays the configuration for the 'Compare diversity indicators' tool (Galaxy Version 0.0.0). The configuration includes fields for 'Input raster' (2: S2A_Subset_test), 'Input header' (1: S2A_Subset_test.hdr), and 'Plots folder zip' (8: S2A_T33NUD_Plots.zip). There are also radio button options for 'Do you want to compute beta diversity (needs multiple locations)?' (Yes selected) and 'Do you want to do a PCA or a SPCA?' (SPCA selected). A blue 'Execute' button is visible. Below the configuration, there is a description of the tool: 'Process satellite remote sensing data to produce biodiversity indicators'. The 'What it does' section explains that the tool is based on a method by Féret and Asner (2014) for tropical forest diversity mapping. The 'Input description' section states that the tool expects an ENVI HDR image with BIL interleave.

Compare diversity indicators



Prosail

Spectral indices

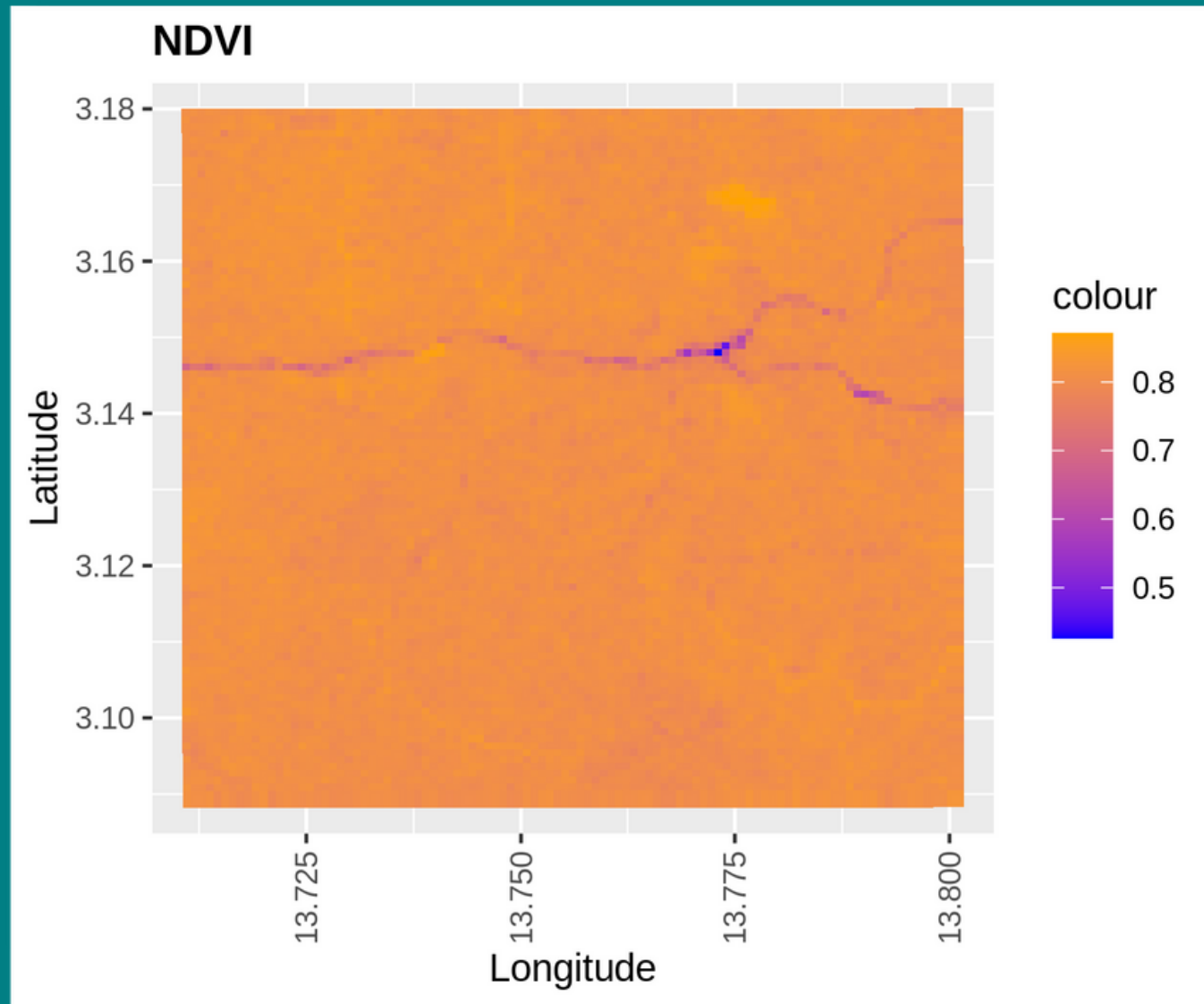
Spectral indices to know more about the well being of the vegetation.

Input

The 2 envi bil files from the first tool

Output

- A tabular on the indice you chose to compute (NDVI, CCCI, ...)
- Graphs on the indices
- The raster layer in envi bil format



S2A_Subset_test	longitude	latitude	NDVI
234.29	13.7115065213769	3.17883791054367	0.809475381273922
233.06	13.7124063428304	3.17883903826377	0.814533326859802
231.57	13.7133061646046	3.17884016519658	0.808063813546791
228.73	13.7142059866993	3.17884129134209	0.814678963045019
230.6	13.7151058091144	3.17884241670032	0.812649359747791
228.85	13.7160056318495	3.17884354127124	0.806605795437084
229.59	13.7169054549044	3.17884466505487	0.810899033966968



Galaxy Configured by Planemo

Workflow Visualize Bibliothèque de données Admin Aide Utilisateur

Tools

search tools

Upload Data

Upload File from your computer

Compute biodiversity indices from remote sensing data

Compare diversity indicators with remote sensing data

Preprocess sentinel 2 data read, crop, resample and write it as a raster stack.

Compute a PCA raster from remote sensing data

Compute spectral indices as NDVI from remote sensing data

Validate ISO 19139 metadata documents from files or Urls.

Map diversity from remote sensing data

WORKFLOWS

All workflows

Compute spectral indices as NDVI from remote sensing data (Galaxy Version 0.0.0)

In which format are your data ?

You already have the files ENVI BIL

You can directly take the Reflectance output from the tool 'Preprocessing sentinel 2 data'

Input raster

2: S2A_Subset_test

Input raster header

1: S2A_Subset_test.hdr

Input the type of indice you want

Select/Unselect all

ARI1 ARI2 ARVI BAI BAIS2 CCCI CHL RE CRI1 CRI2 EVI EVI2 GRVI1 GNDVI IRECI LAI SAVI MCARI mNDVI705 MSAVI2 MSI mSR705 MTCI nBR_RAW NDI_45 NDII NDSI NDVI NDVI_G NDVI705 NDWI NDWI1 NDWI2 PSRI PSRI_NIR RE_NDVI RE_NDWI S2REP SAVI SIPI SR CR_SWIR

Do you want the raster layer of the indice as an output ?

No Yes

Execute

Computes biodiversity spectral indices from satellite remote sensing data

What it does

This tool estimates vegetation properties (leaf and canopy) from sensor measurements. You can choose which vegetation property you want to compute.

Input description

It expects an image file as input, with a specific data format. ENVI HDR image with BIL interleave required. The image is an ENVI raster including :

- A binary file (which has no extension here).
- A header file (with .hdr extension).

The header file is a text file including all necessary metadata which can be read with a text editor. It includes image dimensions, projection, and the name and central wavelength for each spectral band.

How does it look like ?

Sentinel 2 workflow

The screenshot displays the Galaxy workflow editor interface. The main workspace shows a workflow titled "Sentinel 2A biodiversity" with the following steps:

- S2 Data from SciHub**: output (data)
- Preprocessing sentinel 2 data**: Input data, Reflectance (input), Cloud mask (input), Mission source (csv)
- Processing remote sensing data**: Input raster, Input raster header, Input selected components, Plots folder zip, Alpha (input), plots (input), Shannon
- Compute spectral indices**: Input raster, Input raster header, plots (input), ARI1 (tabular)
- Compute biodiversity indices**: Input raster, Input raster header, plots_png (input), Biodiversity indices (tabular)

The right-hand panel contains metadata for the workflow:

- Name**: Sentinel 2A biodiversity
- Version**: 2: Jul 7th 2022, 6 steps
- Annotation**: Sentinel 2A biodiversity workflow
- License**: Specify a license for this workflow.
- Creator**: Add a new creator - either a person or an organization.
- Tags**: biodiversity, gis

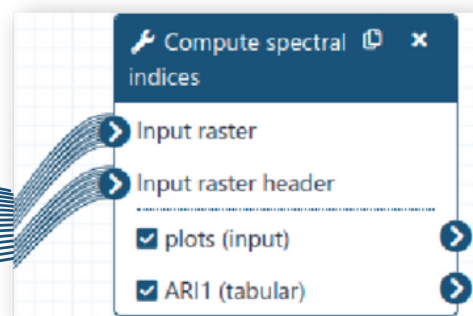
A tutorial is being developed to explained this workflow for biodiversity



SRS - EBV

Satellite remote sensing essential biodiversity variable

Add some of GEO BON EBV to Sentinel 2 workflow which are similar to spectral indices tool



Exemples

- Canopy chlorophyll content
- Vegetation active period
- Red list ecosystem

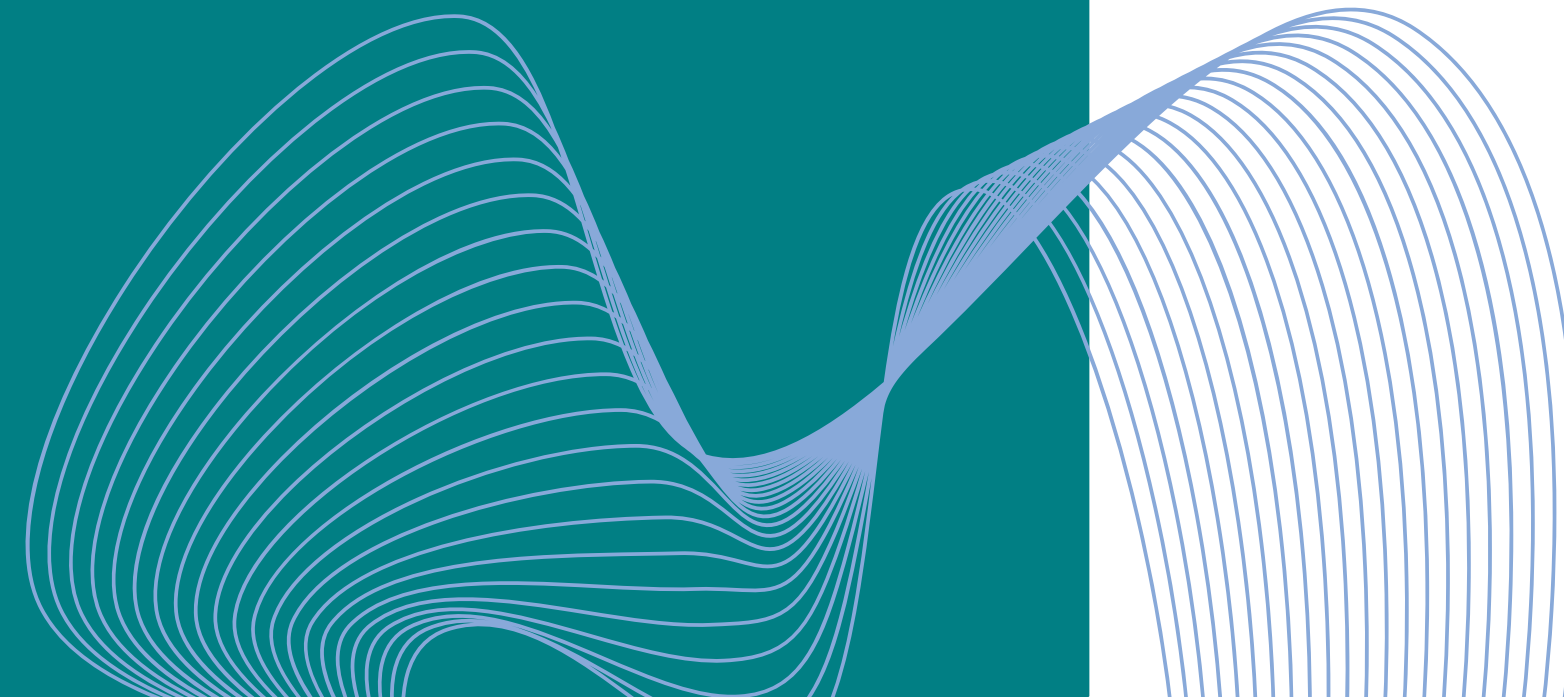


Sentinel 2 workflow for biodiversity indicators

Still in peer review should be available very soon with its
tutorial

Biodiversity data exploration workflow

A fully operational workflow available on Galaxy-
Ecology



Exploration steps

01 Get and clean in-situ data

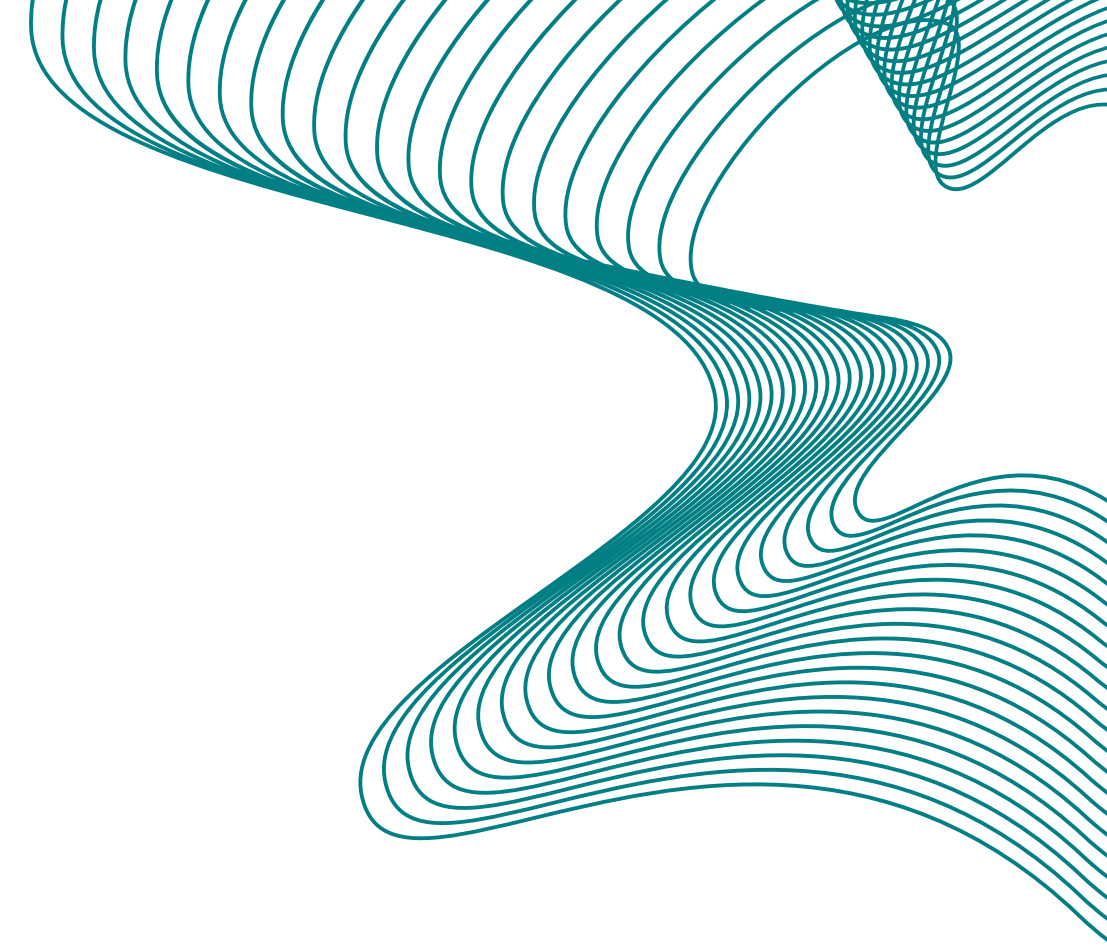
02 Presence, absence and abundance

03 Statistics on presence-absence

04 Variables explorations

05 Homoscedasticity and normality

06 LCBD



Biodiversity

In situ data

Spatial

Remote sensing data





Thank you !