



BIODIVERSITY
BUILDING
BLOCKS FOR
POLICY

Empowering Biodiversity Monitoring through Data Cubes

Techniques & Applications for Open Science

Maarten Trekels & Sandra MacFadyen



Funded by
the European Union



**Meise
Botanic Garden**

NITheCS
National Institute for
Theoretical and Computational Sciences



Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

2, 9, 16, 23 Oct 2024
Online
NITheCS mini-school

Introductions

Biodiversity Monitoring and Data Cubes

Who are we

Online: Wednesdays - 2, 9, 16 & 23 October 2024 @ 14h00-15h00 SAST



Sandra MacFadyen is a research fellow with the [Mathematical Biosciences Lab](#) at Stellenbosch University (SU). She earned a joint PhD in Botany from SU and Landscape Ecology from Vrije University Amsterdam in 2018, alongside an MSc in Geographic Information Science and postgraduate degrees in Nature Conservation. Sandra is a core member of the EU-funded [Biodiversity Building Blocks for Policy](#) programme, a NITheCS Associate, and a principal investigator in a NITheCS research programme. Based in the Kruger National Park with the [SANParks Savanna Science Unit](#), she specialises in macroscale ecosystem dynamics and applied spatial statistics for biodiversity conservation. Through her roles at BioMath and NITheCS, she focuses on advancing novel mathematical models to tackle critical ecological challenges.



Maarten Trekels is a Biodiversity Data Scientist and project coordinator at [Meise Botanic Garden](#), Belgium. With a background in Physics and industry experience in aerospace and medical fields, he shifted to biodiversity data through developing standards for the Biodiversity Information Standards organisation under EU-funded projects related to the [DiSSCo](#) research infrastructure. He currently coordinates a local implementation of the research infrastructure in Flanders and contributes to FAIR data strategies and interdomain interoperability as a member of the Research Data Alliance. Maarten is also a work package leader in the EU-funded Biodiversity Building Blocks for Policy programme and is pursuing a PhD at SU, focusing on community ecology.

Biodiversity Monitoring using Data Cubes

Techniques and Applications for Open Science

Mini-school Outline

Online: Wednesdays - 2, 9, 16 & 23 October 2024 @ 14h00-15h00 SAST

- **Session 1: Introduction** (Sandra and Maarten)
 - A general introduction to **Biodiversity** and **Data Cubes**
 - Sources of Biodiversity **Data** and **Data Standards**
 - Data **publication** and **FAIR** data publishing
 - Homework - List of free accounts to setup before next, more hands-on, sessions
- **Session 2: Building Data Cubes with GBIF** (Maarten)
 - Introduction to GBIF data
 - A Hands-On Guide for Biodiversity Monitoring
- **Session 3: Ecological Modelling with Data Cubes** (Sandra)
 - Create your own Data Cube using GBIF, Google Colab and Earth Engine
 - Visualise results on a interactive map
 - Analyse patterns of species diversity related to environmental variables, like climate
- **Session 4: Open Science** (Maarten and Sandra)
 - Sharing and Disseminating Results
 - Hands-on with *Github* and *Zenodo*
 - Mini-school Wrap Up and what would you like to see next?



BIODIVERSITY
BUILDING
BLOCKS FOR
POLICY

LECTURE 3: ANALYSING DATA CUBES

A practical guide - Techniques and Tools for
Ecological Modelling

Sandra MacFadyen



Funded by
the European Union



**Meise
Botanic Garden**

NITheCS
National Institute for
Theoretical and Computational Sciences



Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

16 October 2024
Online
NITheCS mini-school

The background of the slide is a close-up photograph of a large number of bees, likely honeybees, clustered together on a light-colored, textured surface. The bees are in various positions, some facing the camera and others with their backs to it. The image is overlaid with a semi-transparent green filter, which makes the bees appear in shades of green and yellow. The text is centered over this background.

Analysing Data Cubes

Techniques and Tools for Ecological Modelling

Biodiversity Monitoring using Data Cubes

Techniques and Applications for Open Science

Links to today's material

- Colab code:

https://github.com/nithecs-biomath/mini-schools/blob/main/cubes_prac_3.ipynb

Techniques and Applications for Open Science

Jupyter Notebook in Colab

cubes_prac_3_v3.ipynb ☆

File Edit View Insert Runtime Tools Help Last edited on October 16

Table of contents

- Ecological Modelling
 - System setup
 - Mount Google Drive
 - Import missing or update packages
 - Import, authenticate and activate Google Earth Engine (GEE)
 - Import the rest of the packages
 - Some packages need to be installed each time when using Colab
 - Access data
 - Species occurrence records
 - Local, customised or other GitHub data
 - GBIF data
 - GBIF API through pygbif
 - Design function - fetch_gbif_data
 - Run function - fetch_gbif_data
 - Zip URL
 - Design function - load_from_url_zip
 - Run function - load_from_url_zip
 - Environmental data
 - Setup some parameters
 - Administrative boundaries
 - Long-term rainfall
 - Long-term surface water trends
 - Plot environmental data using geemap
 - Process data
 - Species richness data cube
 - Environmental data cube

Ecological Modelling

Techniques and Tools for Ecological Modelling

In this lecture, participants will explore advanced analytical techniques and tools for ecological modelling using data cubes, with a focus on understanding climate-related impacts and adaptation strategies (SDG 13: Climate Action).

Let's walk through an example of how to use species occurrence data (e.g., from the Global Biodiversity Information Facility, GBIF) combined with climate data in Google Earth Engine (GEE) to understand climate-related impacts on species distributions. We'll also use Google Colab as our computational environment. This example will focus on visualizing the relationship between species occurrences and climate variables, and then assessing potential changes over time.

System setup

Mount Google Drive

In Google Colab, using `drive.mount` allows you to access files for is especially useful when working with larger datasets or pre-existing files that are accessed and manipulated for ecological or biodiversity modeling. This ensures that these resources are available without needing to upload them manually each time.

Follow the prompts to connect your Google Account account.

```
[ ] # Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly

Import missing or update packages

It may be necessary to install specific packages using `!pip` because the Colab environment. By using `!pip`, we can install any additional libraries or dependencies that are not included in the Colab environment. This ensures that we have access to the latest tools, specialized libraries, or dependencies that are necessary for the analysis. Once these dependencies are installed, we can then proceed with the analysis using the Colab environment.

```
>> Run function - load_from_url_zip
```

```
# URL path
url_zip = 'https://api.gbif.org/v1/occurrence/download/request/0038969-240906103802322.zip'

# Run function
df_url = load_from_url_zip(url_zip)

# Specify the columns to keep
df_url = df_url[['year', 'month', 'family', 'speciesKey', 'species', 'decimallatitude', 'decimallongitude']]
df_url.head(4)
```

	year	month	family	speciesKey	species	decimallatitude	decimallongitude
0	2013	10	Anatidae	5232465.0	Plectropterus gambensis	-25.162979	31.617205
1	2011	12	Anatidae	5232465.0	Plectropterus gambensis	-25.559278	30.059526
2	2010	12	Anatidae	5232465.0	Plectropterus gambensis	-34.062473	18.496857
3	2011	12	Anatidae	5232465.0	Plectropterus gambensis	-26.026592	27.855534

Code cell output actions

- All layers on/off
- OpenStreetMap...
- Total Water Occur...
- Monthly Mean Rai...

Species Richness Change Over Time

Biodiversity Monitoring using Data Cubes

Techniques and Applications for Open Science

Colab - Mount GDrive

Follow the prompts to connect your Google Account account.

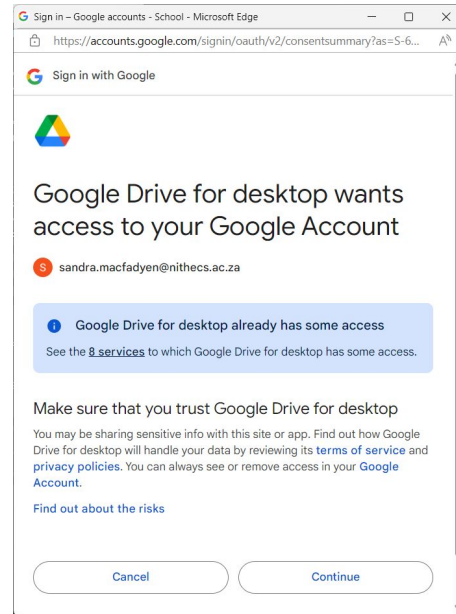
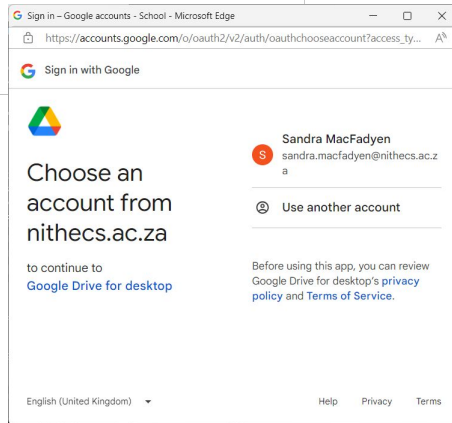
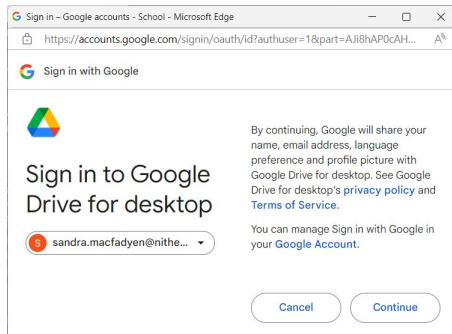
```
✓ 10s # Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

Permit this notebook to access your Google Drive files?

This notebook is requesting access to your Google Drive files. Granting access to Google Drive will permit code executed in the notebook to modify files in your Google Drive. Make sure to review notebook code prior to allowing this access.

No thanks [Connect to Google Drive](#)



Biodiversity Monitoring using Data Cubes

Techniques and Applications for Open Science

Colab - Import, Authenticate and Initialize Google Earth Engine

- ! The first time you do this you'll be prompted to add a 'token key'. `ee.Authenticate()` should initiate pop-up tabs with prompts to get to retrieve your unique 'token key'. Copy and paste this into the cell that will appear in your jupyter notebook !
- The next time you `ee.Initialize()`, the following prompts should appear:

Follow the prompts to connect your GEE account.

```
[ ] # Import, authenticate and activate Earth Engine (EE)
import ee

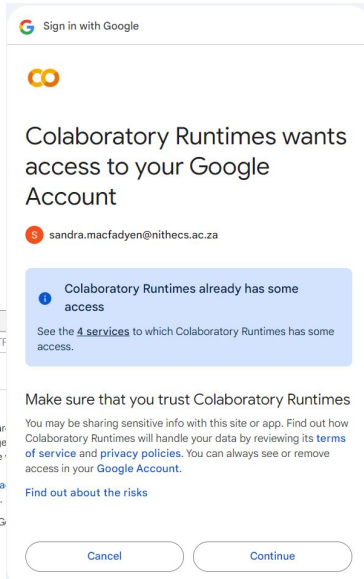
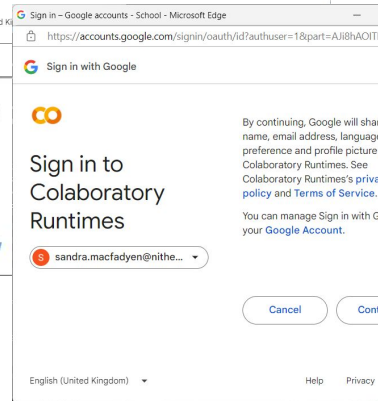
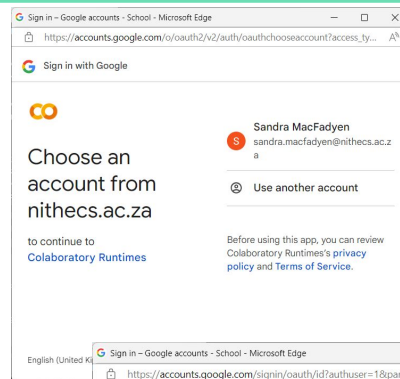
# Trigger the authentication flow.
ee.Authenticate()

# Initialize the library.
ee.Initialize(project='ee-biodatacube') # Change according to your own project name
```

Allow this notebook to access your Google credentials?

This will allow code executed in this notebook to access your Google Drive and Google Cloud data. Review the code in this notebook prior to allowing access.

No thanks Allow



Analysing Data Cubes

Techniques and Tools for Ecological Modelling

Questions?



END OF DAY 3

Thank you!

Maarten Trekels | Sandra MacFadyen

maarten.trekels@plantentuinmeise.be | macfadyen@sun.ac.za



b-cubed.eu



@BCubedProject



B-Cubed Project



B-Cubed
Newsletter