

Open Data Science: Reproducible and Sharable Data Analytical Work Flows in Ecology Impulsvortrag

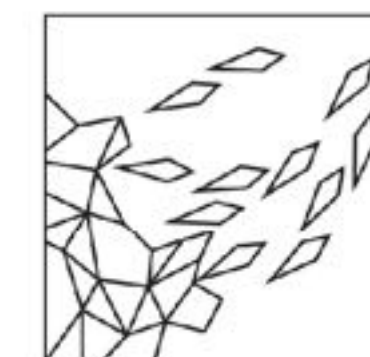
Urs Kalbitzer



MAX PLANCK INSTITUTE
OF ANIMAL BEHAVIOR



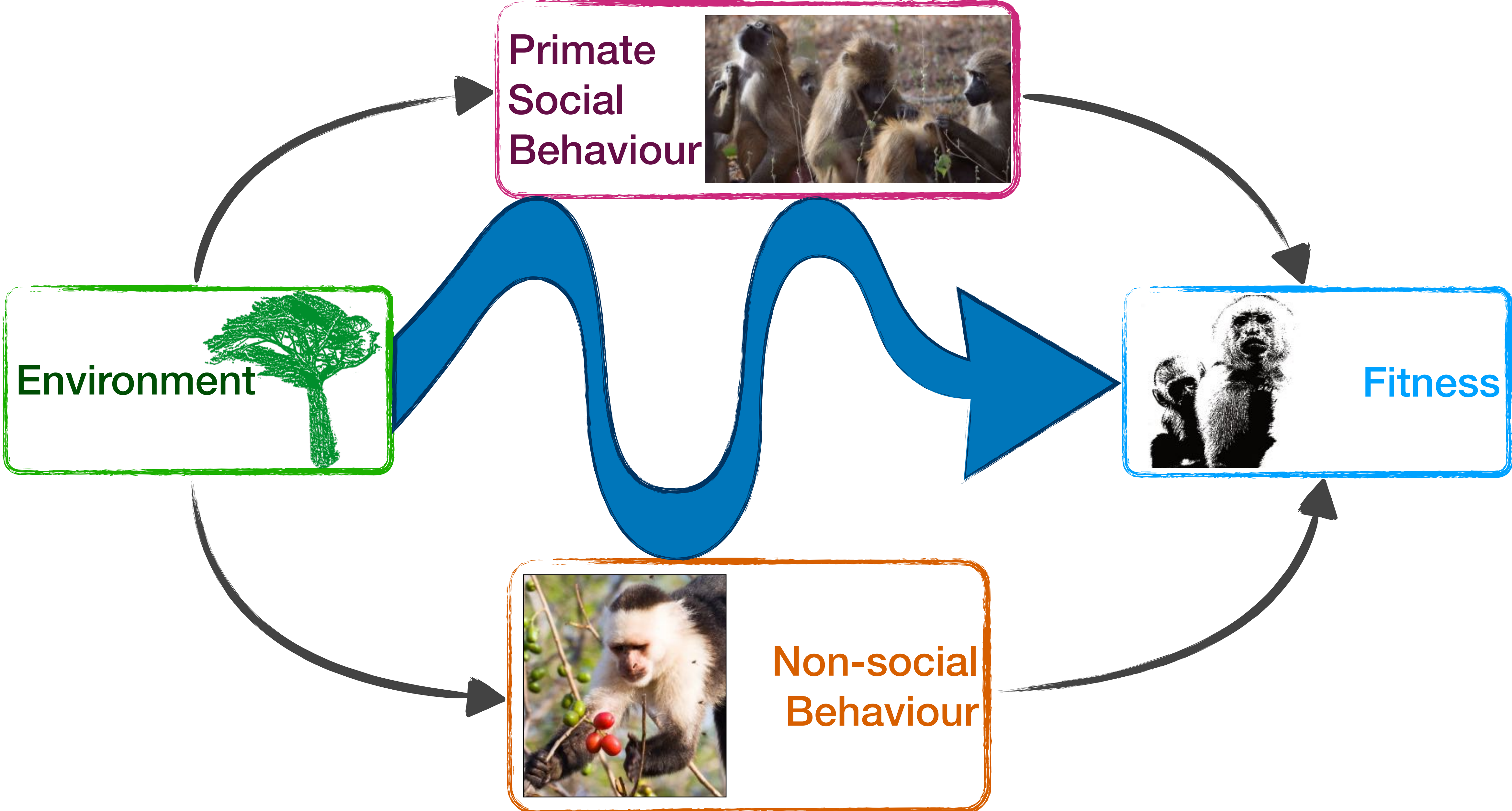
Centre for the Advanced Study
of Collective Behaviour



Universität
Konstanz



My Research Interests



**Focus here: How are we doing
science?**

***Reproducible and sharable data analytical
work flows (Open Data Science)***

**Essential to enhance transparency and
reusability of scientific findings**

Background

13-week workshop series on *Open Data Science in the Department for the Ecology of Animal Societies* (Prof. Meg Crofoot at MPI-AB & Uni Konstanz):

Create common working practices for collaborative, reproducible, and rigorous science¹

- Values and tools to facilitate **collaborations** with others and the *future-self*
 - How to **collect, store, and share data** and other materials?
 - How to **document your work, data processing steps, code** (including version control), **and preliminary results?**
 - Requirements for **code sharing and collaborations**

¹Based on Lowndes et al. (2019): **Supercharge your research: A ten-week plan for open data science.** *Nature*.

Open Data Science

GitHub Wiki: https://github.com/livingingroups/ods_wiki/wiki

Willkommen!

Welcome to the MPI-AB EoAS Open Data Science wiki!

Here you can find links to protocols, resources and tutorials regarding our department's goals of reproducible research, open data science, and science reform. This is the collaborative outcome of the department's [12 sessions of ODS meetings](#) that occurred in Winter and Spring 2021.

1. [Tips, Tools and Requirements for Project Management and Organization](#)
2. [Collecting, Managing, and Backing Up Data](#)
3. [How to access and use the server](#)
4. [Version Control and Git](#)
5. [Best Practices for Data and Code Management](#)
6. [Resources for Coding, Reproducible Research and Open Data Science](#)
7. [Research Notebooks](#)
8. [Internal Code Review](#)

Department Requirements

► Details

Department Aims

► Details



▼ Pages 8

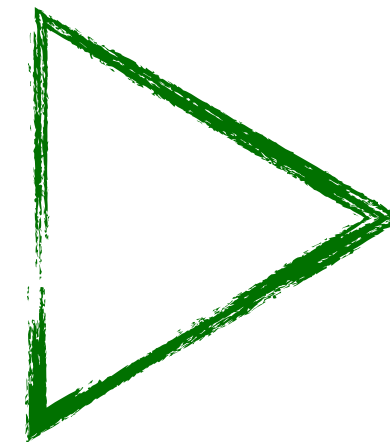
Find a Page...

▼ Home

- Willkommen!
- Department Requirements
 - Data Management Plan
 - Data Catalog Entry Form
 - Upload Raw Data to server
 - Upload Processed Data to Server
 - Project or Research notebooks
 - Internal Code Review and Publication Documentation
- Department Aims
 - 1. Before Data are Collected
 - 2. During Data Processing
 - 3. During and After Data Analysis
- [Best Practices For Data and Code Management](#)
- [Department Aims](#)
- [EAS Data Management Protocol](#)
- [EAS Data Server Instructions](#)
- [Internal Code Review](#)
- [Research notebooks](#)
- [Version control and Git](#)

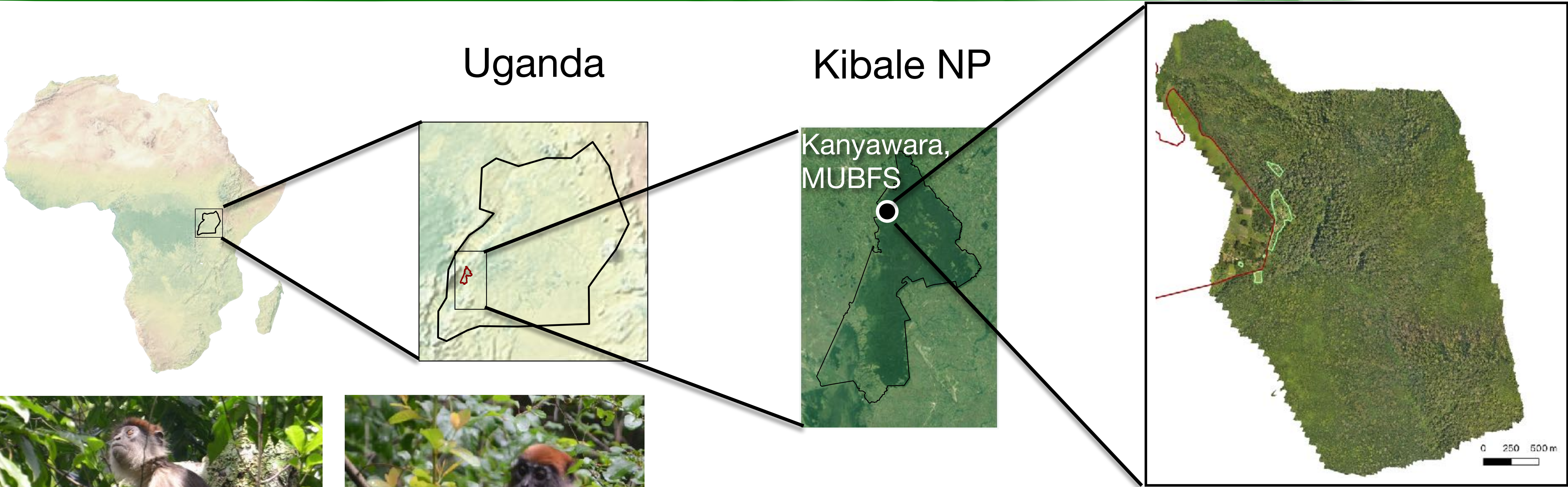
Open Data Science - Our Aims

- 1. *Before data are collected:*** Collect valuable data that are accurate and reproducible
- 2. *During data processing:*** Process data carefully, deliberately, and reproducibly
- 3. *During and after data analysis:*** Analyse data clearly, honestly, and reproducibly
 - ***Clear structure***
 - ***Commenting code and using consistent style***
 - ***Version control***
 - ***Share code with collaborators***



Git/GitHub

Long-term project in Kibale National Park, Uganda



How are climatic factors linked to primate food availability?

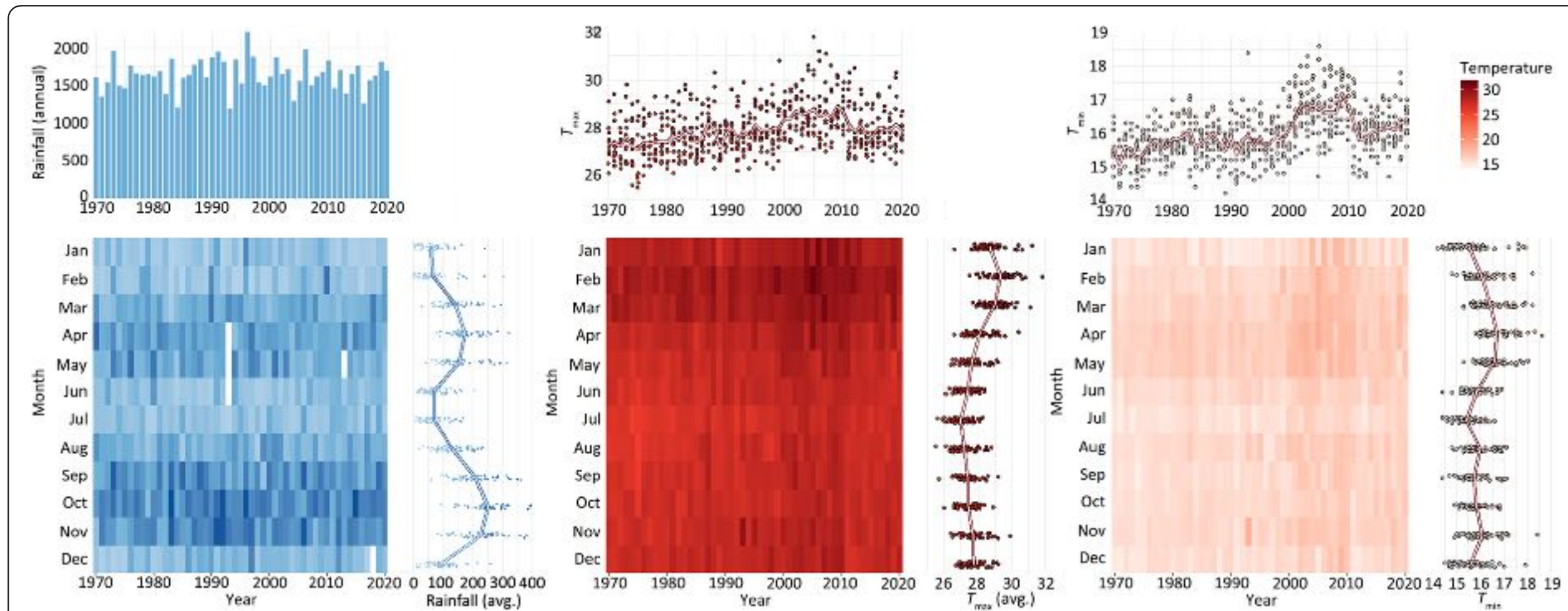


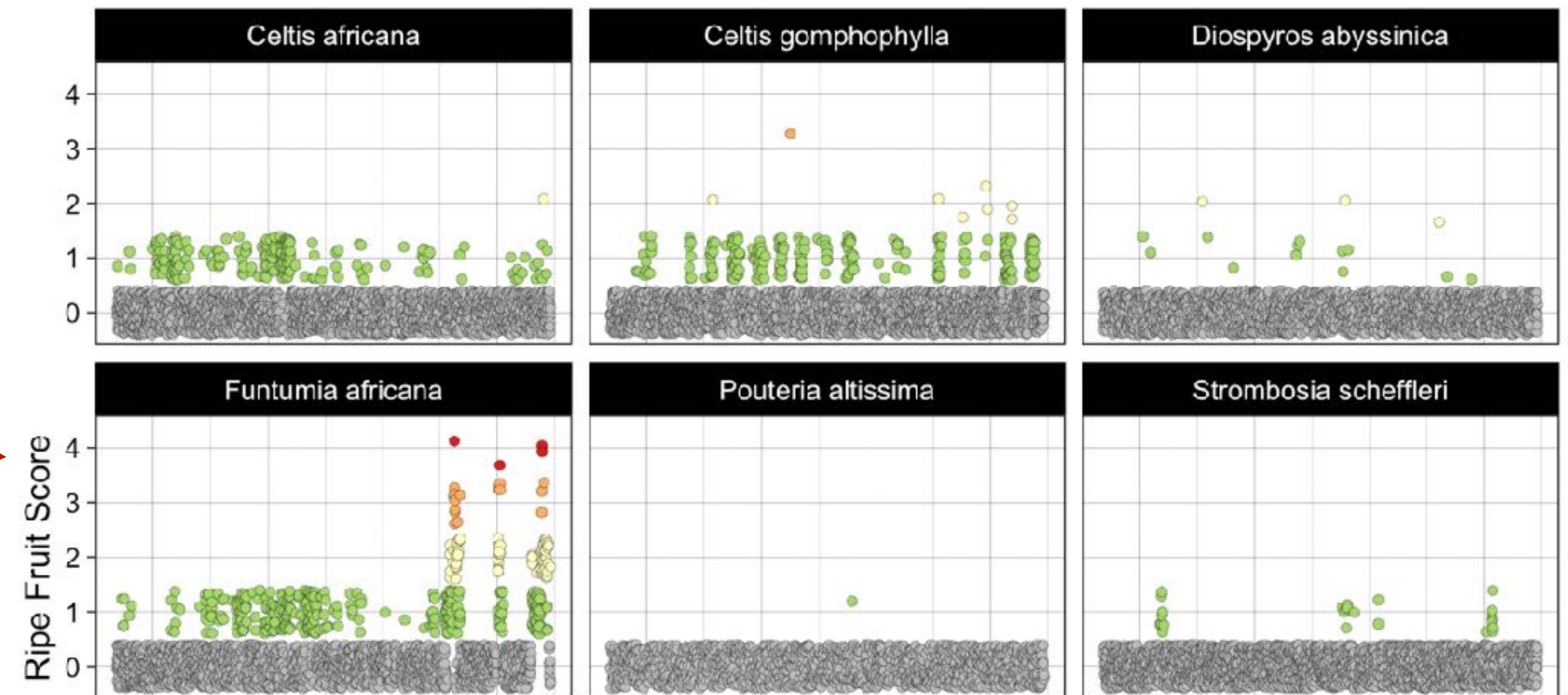
Fig. 1 Patterns of rainfall, maximum temperature ($^{\circ}\text{C}$, T_{max}) and minimum temperature ($^{\circ}\text{C}$, T_{min}) between 1970 and 2020 for the area near Makerere University Biological Field Station in Kibale National Park, Uganda. For details, see text

Chapman, C. A., ... & Kalbitzer, U. (2021). A 40-year evaluation of drivers of African rainforest change. *Forest Ecosystems*, 8(1), 66. <https://doi.org/10.1186/s40663-021-00343-7>

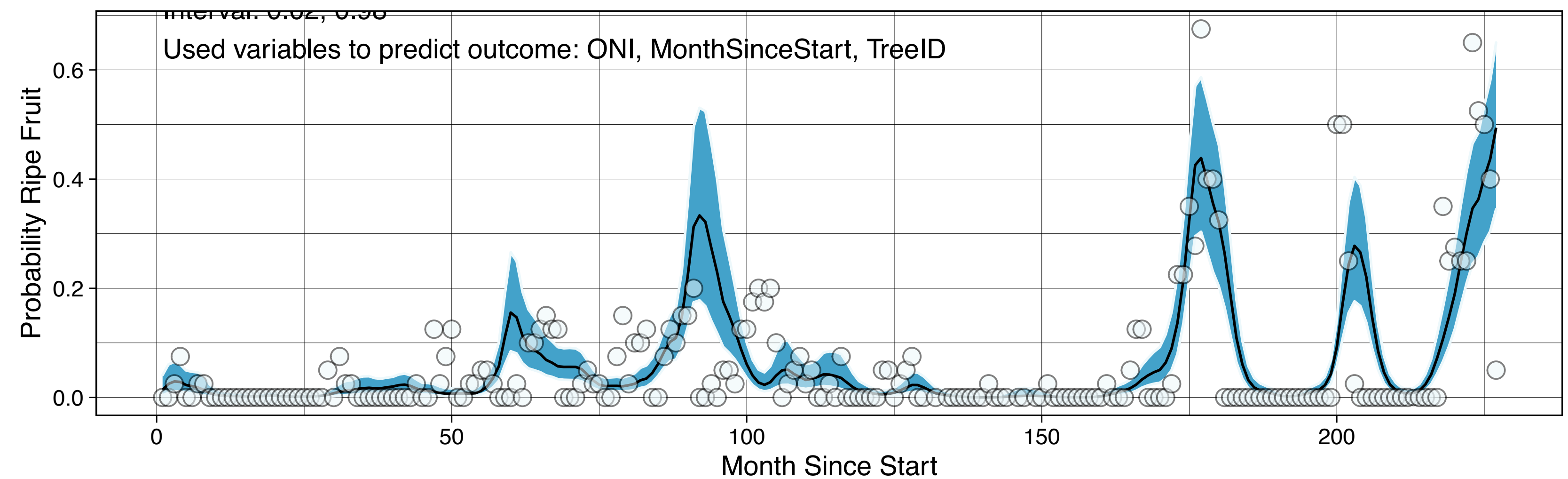
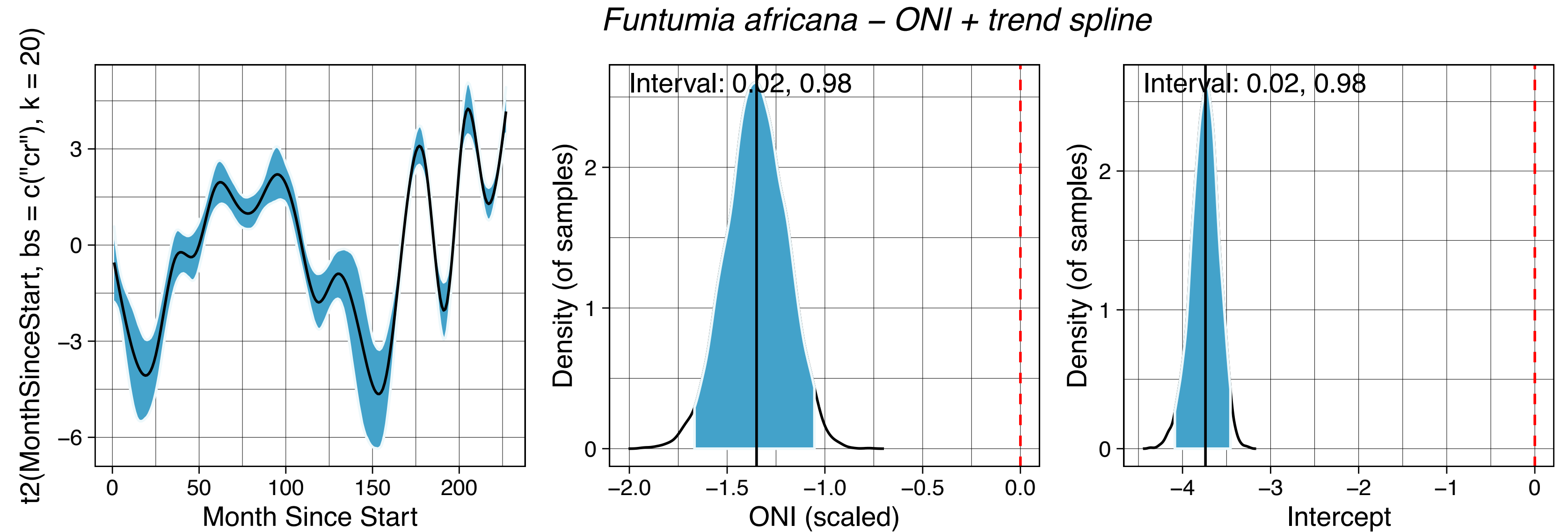
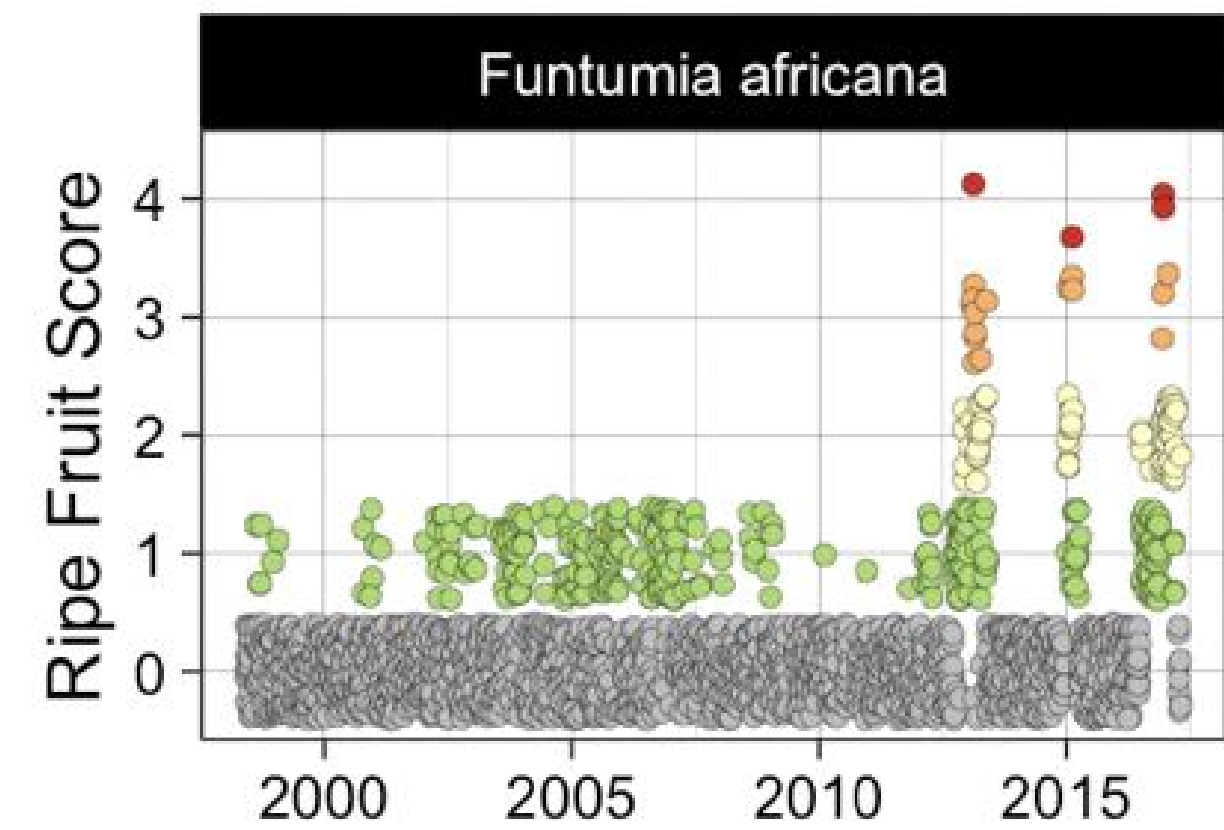
Plant Productivity

Climate

?



Example: *Funtumia africana*



Project Structure

The screenshot shows the GitHub repository page for 'urskalbitzer / Phenology_Fruit_40y_R'. The repository is private and has 1 branch and 0 tags. The commit history shows several updates to the webpage and data processing. The file list includes folders for raw data, processed data, documentation, model results, notebooks, plots, scripts, and webpage source, along with configuration files like .DS_Store, .gitignore, and a Rproj file.

Annotations:

- "Raw" data:** Points to the `.Rproj.user` folder.
- Data processed within project:** Points to the `data_input` and `data_intermediate` folders.
- All scripts required for the analysis:** Points to the `scripts` folder.
- Bookdown webpage:** Points to the `webpage_source` folder.

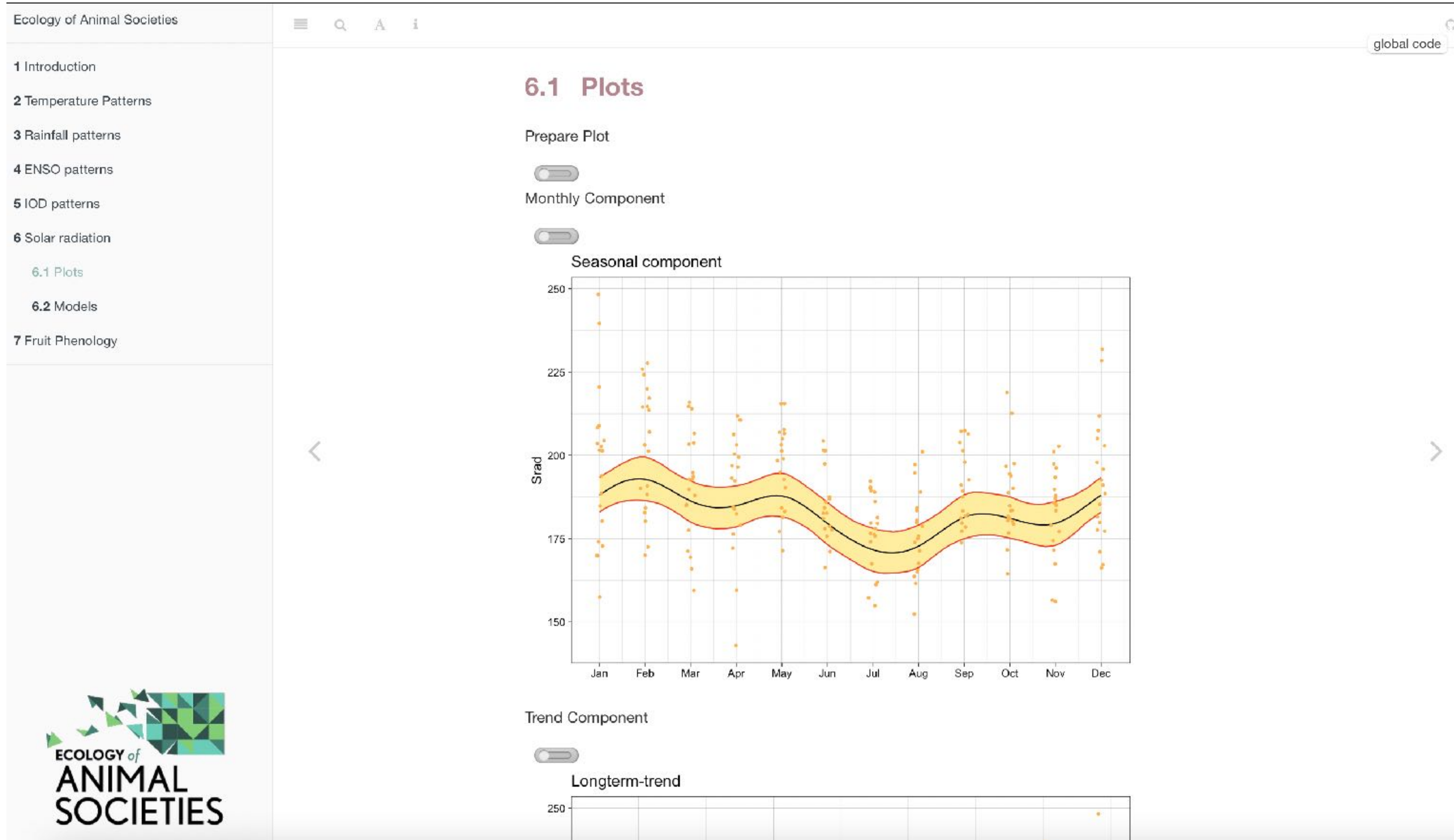
File/Folder	Commit Message	Time Ago
<code>.Rproj.user</code>	Updates to webpage	2 days ago
<code>data_input</code>	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
<code>data_intermediate</code>	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
<code>docs</code>	Updates to webpage	2 days ago
<code>model_results</code>	Wrote functions to run models on all species	4 months ago
<code>notebooks</code>	Update phenodata, worked on models	5 months ago
<code>plots</code>	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
<code>scripts</code>	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
<code>webpage_source</code>	Updates to webpage	2 days ago
<code>.DS_Store</code>	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
<code>.gitignore</code>	Wrote functions to run models on all species	4 months ago
<code>Pheno_Fruit_40y.Rproj</code>	First upload	8 months ago

Repository Metadata:

- Repository URL: urskalbitzer.github.io/phenology_fruit...
- Readme: [Readme](#)
- Stars: 0
- Watching: 1
- Forks: 0
- Releases: No releases published. [Create a new release](#)
- Packages: No packages published. [Publish your first package](#)
- Environments: 1 (github-pages is Active)

Sharable Webpage

https://urskalbitzer.github.io/Phenology_Fruit_40y_R/



Structure of Code

main ▾ Phenology_Fruit_40y_R / scripts /

Go to file Add file ▾ ...

urskalbitzer Added ENSO and IOD models, updated all plots and bookdown ... ✓ 3cd7eaa 2 days ago History

..		
plotting	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
process_data	Added SRad, updated webpage	4 months ago
run_models	Added ENSO and IOD models, updated all plots and bookdown	2 days ago
temp	Finalized 1st function draft create pheno plot data again	2 months ago
.DS_Store	Worked on models for c africana, especially monthly models	5 months ago
Longterm_Rain_Temp_Plot.R	First upload	8 months ago
runall.R	Added ENSO and IOD models, updated all plots and bookdown	2 days ago

```
544 lines (489 sloc) 22.1 KB Raw Blame
```

```
1 # Project: Analyzing 40 years of fruit phenology in Kibale #
2
3 rm(list = ls())
4 library(dplyr)
5 library(tidyr)
6 library(stringr)
7 library(sf)
8 library(stars)
9
10 # ++++++
11 ### Get and process phenology data
12 # ++++++
13 # pheno.tss <- read.csv(here::here("data_input/Phenodata_processed/struhsaker_1970_1985.csv"))
14 # pheno.pt <- read.csv(here::here("data_input/Phenodata_processed/phencrtrail_1998_until_now_processed.csv"))
15 # pheno.kan <- read.csv(here::here("data_input/Phenodata_processed/KAN_1998_1998_processed.csv"))
16 #
17 # # For each dataset:
18 #
19 # # Create unique Tag-IDs
20 # # Check if species always the same for a single TAG-ID
21 # # Check DBID?
22 #
```


Next Steps

- Finalize Project
 - Code review by collaborators
 - Publish:
 - Results - as journal paper
 - Code - As repository? As Supplementary Material?
 - Processed data set - As repository? As data paper?
- ➡ Will allow to reproduce the analysis

Thank You!

Comments or questions?

urskalbitzer.github.io
urs.kalbitzer@uni-konstanz.de
ukalbitzer@ab.mpg.de