

# Environmental Systems Data Science

Loïc Pellissier, Joshua Payne, Benjamin Stocker

2020-06-24



# Contents

<b>1</b>	<b>Primers</b>	<b>5</b>
1.1	Exercises . . . . .	5
1.2	How to use Bookdown and Markdown . . . . .	7
<b>2</b>	<b>Big Data for environmental sciences</b>	<b>9</b>
2.1	Exercises . . . . .	9
<b>3</b>	<b>Data scraping and wrangling</b>	<b>11</b>
3.1	Exercises . . . . .	11
<b>4</b>	<b>Model-data fusion</b>	<b>13</b>
4.1	Exercises . . . . .	13
<b>5</b>	<b>Supervised learning</b>	<b>15</b>
<b>6</b>	<b>Unsupervised learning</b>	<b>17</b>



# Chapter 1

## Primers

(Youtube videos can be embedded, according to this, but it didn't work for me.)

### 1.1 Exercises

#### 1.1.1 Overview

- **Input:** Reads for the first time the dataset 1 of two (Swiss) FLUXNET sites at high temporal resolution with 3-4 measures.
- **Output:** Learned about the data structure and the capacity to explore the content of the dataframe and visualize it so that it can be analysed in the next session.
- **Data:** Dataset 1 First hands-on with one of two (Swiss) FLUXNET sites at high temporal resolution with 3-4 measures with high-resolution fluxnet dataset

#### Contents

- Content and operations:
- Managing the workspace and data
- Basic operations in R
- RStudio, debugging
- Reading a table into a data frame
- Objects, data frames, lists
- Applying a function, loop, etc.
- Simple plotting
- Making workflow reproducible
  - RMarkdown

- Coding
  - Best practices
  - Modularity
- Version control with git (for dummies: GitHub, fork, clone, commit, push, pull, branch)

### 1.1.2 Reading data into R

Load libraries.

```
library(readr)
library(tidyverse)
```

Read data from Laegern site.

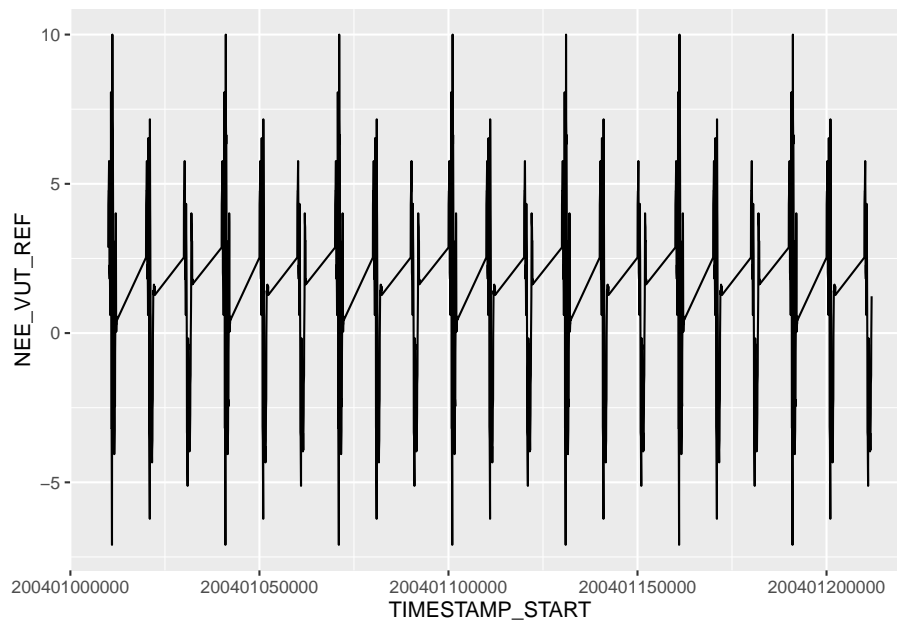
```
df_fluxes <- read_csv("./data/FLX_CH-Lae_FLUXNET2015_FULLSET_HH_2004-2014_1-3.csv", )
```

```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

Plot something.

```
ggplot(data = slice(df_fluxes, 1:1000), aes(x = TIMESTAMP_START, NEE_VUT_REF)) +
  geom_line()
```



## 1.2 How to use Bookdown and Markdown

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation  $a^2 + b^2 = c^2$ .

The **bookdown** package can be installed from CRAN or Github: nothing

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

If you want to generate PDF output, you will need to install LaTeX. For R Markdown users who have not installed LaTeX before, we recommend that you install TinyTeX (<https://yihui.name/tinytex/>):

```
install.packages('tinytex')
tinytex::install_tinytex() # install TinyTeX
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-





## Chapter 2

# Big Data for environmental sciences

### 2.1 Exercises

We have time series data obtained from multiple sites. A common task is to complement our data with information about the sites (meta data), or with data from other sources and analyse it using this additional data. In this exercise, you will learn to...

#### 2.1.1 Overview

- **Input:** From the last session they learned how to look at two flux tower site of dataset 1.
- **Output:** They have gotten familiar with the dataset 2, which contains multiple site

#### Content

- Meta data
- Combining data: For our set of FLUXNET sites, collect data from other sources (global fields, remote sensing)
- Finding, accessing, downloading, and reading environmental data from the web, web scraping
- MODIS, Landsat, Google Earth Engine, Worldclim, Chelsa, CRU, (some re-analysis), CMIP5 from Urs, HWSD, ETPO1, Biome classification, IGBP veg types, Koeppen-Geiger climate, .....

- Merging, nesting
- Formats: geo data: raster, NetCDF, CSV, shapefile
- bash, cron, wget, grep, system, RegExp

### 2.1.2 Read multiple files

Find all daily time series data in the `./data` directory. Files are identified here by their name, which contains the pattern `"DD"` (for ‘daily’).

```
filelist <- list.files("./data", pattern = "DD")
```

This returns 19 files for 19 sites. We can read them in at once using a simple loop. Here, we are creating a list of data frames of length 19.

```
library(readr)
list_df <- list()
for (ifil in filelist){
  list_df[[ifil]] <- read_csv(paste0("./data/", ifil))
}
```

In *tidyverse*, this could be done on one line by:

```
library(purrr)
list_df <- purrr::map(as.list(filelist), ~read_csv(paste0("./data/", .)))

## This returns a unnamed list. Let's add names as done above.
names(list_df) <- filelist
```

It may be impractical to have the different dataframes as elements of a list. In fact, the data frames read in here each have similar shapes. I.e., they share the same columns (but differ by their number of rows, and of course, by their data values). This suggests that we can “stack” each dataframes along rows.

```
library(dplyr)
df_allsites <- bind_rows(list_df, .id = "siteid")
```

This creates one single data frame containing all sites’ data (>90’000 rows), and adds a column named `"siteid"` that is automatically created by using the names of the list elements of `list_df`.

## Chapter 3

# Data scraping and wrangling

### 3.1 Exercises

- **Input:** The students
- **Output:** The student will produce a plant functional trait maps for Europe that can be considered later as a predictor. They have c about Biome classification, IGBP veg types
- **Data:** Dataset 2 in particular the geographic locations for which vegetation types will be produced in combination with plant occurrences (GBIF) and plant trait data (leaf trait - TRY).



## Chapter 4

# Model-data fusion

### 4.1 Exercises

- **Input:** Dataset 1 (half-hourly) read, aggregated to daily values, cleaned and gapfilled
- **Output:** Linear regression ( $GPP \sim PAR$ ); optimized modified model ( $GPP \sim PAR, T, VPD, \text{soil moisture}$ ); Bayesian calibration?
- **Data:** LUE model to FLUXNET GPP



## Chapter 5

# Supervised learning

We have finished a nice book.





## Chapter 6

# Unsupervised learning

We have finished a nice book.