

Land-use data harmonization

Vincent Pellissier

8 November 2018

This report explains the steps taken to correct the land-use data. The masterfile used is the version 1.8, as found on Novembre 8th 2018. The sheet in Grassplot 1.8.xlsx containing the LU information is the sheet 'datasets' and was read and saved as a .rds file to ease the process (faster loading)

```
df <- readRDS(file.path(path_grassplot, 'Grassplot 1.8_Data.rds'))
df <- df %>%
  mutate_at(.vars = c(95:104), funs(ifelse(. %in% c('NA', '[NA]'), NA, .)))
```

The columns 98 (mowing frequency) and 100 (grazing intensity) are renamed:

```
names(df)[c(98,100)] <- c('mowing_frequency', 'grazing_intensity')
```

Correction of the land use column

The file lookup_table_LU.xlsx contains two sheets:

- land_use_detail: lookup table between the column 96 and other column. This needs to be discussed further before I finalize the table.
- global_land_use: lookup table matching the values in column 95 with a new classification. This new classification is a set of binary columns (allowing for mixed land-use) plus one declarative column. Each binary column correspond to a management practice (grazed, mown, burnt, fertilized, abandoned, natural_grassland), the declarative column stores additional information that cannot be put in one of the 6 columns (*i.e* trampled) and that will be used afterward. **NOTE** LAS, P, L were noted as land-use for the dataset PL_C and N was noted as land-use for the dataset RU_J. I could not find what these stand for. I assumed that these plots don't have a land-use but the info is still in the column 'other'.

```
lut <- read_excel(file.path(path_grassplot, "lookup_table_LU.xlsx" ),
  sheet = 'global_land_use')
```

The global_land_use lookup table is used to harmonize the column 95 with the new binary columns:

```
df <- df %>%
  left_join(lut, by = c("Land use (5 standard categories: mown, grazed, abandoned, natural grassland,
```

Matching new binary columns and intensity columns

In this step, we identify datasets with discrepancies between the new binary columns and the matching intensity column to:

- Manually check in the original datasets or publications
- Correct the discrepancies.

Mowing and mowing intensity

Datasets containing plots for which mowing intensity > 0 & mown != 1) (here and after, we refer to the new binary columns)

```
df %>%
  filter(`mown` != 1 & `mowing_frequency` != '0') %>%
  distinct(`Dataset ID`) %>%
  pull()
```

```
## [1] "EU_K" "CZ_J"
```

- EU_K contains **XX** plots classified as natural_grassland which have a mowing frequency. These plots will also be classified as mown
- EU_K contains **XX** plot classified as abandoned which have a mowing frequency. Only the plot with a frequency ≥ 0.2 will be classified as mown (abandoned less than 5 years).

```
df[df$`Dataset ID` == 'EU_K',] %<>%  
  mutate(mown = ifelse(!is.na(mowing_frequency), 1, 0))
```

```
df[df$`Dataset ID` == 'EU_K',] %<>%  
  mutate(mown = ifelse((is.element(mowing_frequency, c('0.03', '0.05')) | is.na(mowing_frequency)), 0, 1))
```

- CZ_J contains plot classified as abandoned which have a mowing frequency $= 0.05$. These will not be classified as mown. (I could not find the original land-use information in CZ_J.xls)

Grazing and Grazing intensity

Datasets containing plots for which grazing intensity > 0 & grazed $\neq 1$)

```
df %>%  
  filter(`grazed` != 1 & `grazing_intensity` != '0') %>%  
  distinct(`Dataset ID`) %>%  
  pull()
```

```
## [1] "IR_A" "PL_D" "TR_B" "UA_G" "EU_K"
```

- IR_A contains 34 plots classified as mown which have a grazing_intensity $= 0.1$. These plots will be classified as grazed (and also remain mown as indicated in the original database IR_A.xls). **Note for Idoia** In the original DB, the grazing intensity is noted as 1, 2, 3. Is it correct that it stands for low, medium and high intensity, and hence translate as 0.1, 0.5, 1 in the master file?

```
df[df$`Dataset ID` == 'IR_A',] %<>%  
  mutate(grazed = ifelse(grazing_intensity != '0', 1, 0))
```

- PL_D contains 39 plots classified as mown which have a grazing_intensity $= 0.5$. These plots will be classified as grazed. (No composition data, so I could not check the original dataset)

```
df[df$`Dataset ID` == 'PL_D' & df$grazing_intensity %in% '0.5',] %<>%  
  mutate(grazed = 1)
```

- TR_B contains 32 plots classified as abandoned which have a grazing_intensity $= 0.1$. These plots will be classified as grazed (no further info could be found in TR_B.xls)

```
df[df$`Dataset ID` == 'TR_B' & df$grazing_intensity %in% '0.1',] %<>%  
  mutate(grazed = 1)
```

- UA_G contains reps. 12, 15, 30 and 3 plots classified as natural_grassland which have a grazing_intensity $=$ 'high', 'low', 'middle' or 'overgrazing'. These plots will be classified as grazed (I could not find the original land-use information in UA_G.xls)

```
df[df$`Dataset ID` == 'UA_G' & df$grazing_intensity %in%  
  c('high', 'low', 'middle', 'overgrazing'),] %<>%  
  mutate(grazed = 1)
```

- EU_K contains 4 plots classified as natural_grassland which have a grazing_intensity $= 0.5$. These plots will be classified as grazed (no further info could be found in EU_K.xls)

```
df[df$`Dataset ID` == 'EU_K' & df$grazing_intensity %in% '0.5',] %<>%  
  mutate(grazed = 1)
```

Matching new and old binary columns

In this step, we identify and correct discrepancies between the newly created and corrected binary column, and the former ones

Mowing

There are:

- 38 plots with `mown == 1` & `Mowing (1/0) == 0`
- 148 plots with `mown == 0` & `Mowing (1/0) == 1`
- 6810 plots with `mown == 1` & `Mowing (1/0) = NA`
- 6 plots with `mown = NA` & `Mowing (1/0) == '?'`

Mowing (1/0)	0	1	<NA>
?	NA	NA	6
0	6217	38	22
1	142	478	NA
NA	150296	6810	16068

Datasets containing plots with `mown == 1` & `Mowing (1/0) == 0` (new vs former column):

```
df %>%  
  filter(`mown` == 1 & `Mowing (1/0)` == '1' )%>%  
  distinct(`Dataset ID`)%>%  
  pull()
```

```
## [1] "AS_A" "IR_A" "PL_A" "PL_D" "TJ_A" "EU_K" "AT_D" "CZ_J"
```

Datasets containing plots with `mown == 0` & `Mowing (1/0) == 1` (new vs former column):

```
df %>%  
  filter(`mown` == 0 & `Mowing (1/0)` == '0' )%>%  
  distinct(`Dataset ID`)%>%  
  pull()
```

```
## [1] "AS_A" "CH_D" "DE_R" "HU_B" "IR_A" "KZ_A" "MN_B" "MN_C" "NP_A" "PL_A"  
## [11] "RU_I" "RU_K" "TJ_A" "TR_B" "UA_H" "UA_I" "UA_J" "EU_K" "AT_D" "CZ_J"
```

Datasets containing plots with `mown == 1` & `Mowing (1/0) = NA` (new vs former column)

```
df %>%  
  filter(`mown` == 1 & is.na(`Mowing (1/0)`))%>%  
  distinct(`Dataset ID`)%>%  
  pull()
```

```
## [1] "AT_A" "BG_A" "CH_A" "CH_B" "CH_C" "CZ_A" "CZ_B" "CZ_C" "CZ_H" "CZ_I"  
## [11] "DE_A" "DE_E" "DE_F" "DE_H" "DE_I" "DE_J" "DE_L" "DE_N" "DE_O" "DE_P"  
## [21] "ES_A" "ES_C" "EU_A" "EU_B" "EU_C" "EU_E" "EU_G" "EU_J" "FR_A" "HR_A"  
## [31] "HU_A" "HU_C" "IT_H" "IT_K" "JP_A" "LV_A" "NL_A" "PL_B" "PL_C" "PL_D"  
## [41] "RO_A" "RO_B" "RS_A" "RU_A" "TJ_A" "UA_C" "UA_D"
```