Report land-use data

Vincent Pellissier

5 November 2018

This report provides information on the structure of the land-use information present in the GrassPlot dataset, along with the code used.

Dataset

The master file (Grassplot 1.8.xlsx as of Nov. 5th 2018) contains five sheets:

- log book: tracking of the modifcations of the masterfile
- Data: list of plot data (part of the datasets) contributed to Grassplot, with info on the plot data (land-use data, environmnental data)
- datasets: list of the datasets contributed to Grassplot, with metadata on the datasets (nested or not, are compositional data present, are there any environmental data...)
- new references: new (?) references added to the data
- Consortium members: list of the consortium members along with info

The sheet containing land-use information at the plot level is the 'Data' sheet. As the last 11 columns (with LU information) have coding error (ex. decimal separator, characters in a numerical column, ...), these columns are read as text columns and will be corrected later. Note that when loading columns with a comma as separator as a text column, the comma is replaced by a period (for an unknown reason...)

```
# dataframe with all the informations at a plot level
df <- read_excel(file.path(path_grassplot, "GrassPlot 1.8.xlsx" ),
                 sheet = 'Data',
                 col_types = rep('guess', 94), rep('text', 11))
dim(df)
```

```
## [1] 180077    104
```

Land-use information

Overview of the information available

There are 11 columns containing potential information about land-use:

```
##  [1] "Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)"
##  [2] "Land use detail (e.g. number of cuts or years since abandonment)"
##  [3] "Mowing (1/0)"
##  [4] "Mowing frequency: cuts per year (2=2cut/yr; 1=1cut/yr; 0.5=1 cut/2 yr or 2 yr abandoned; 0.2=1
##  [5] "Grazing (1/0)"
##  [6] "Grazing intensity (0/1) 1= intensive grazing"
##  [7] "Burning (1/0)"
##  [8] "Ex arable (1/0)"
##  [9] "Ex arable years (years since last ploughing)"
## [10] "Fertilized (1/0)"
```

Some of these columns contains hardcoded NAs (*i.e.* as character chain, 'NA' or '[NA]') which can be replaced by proper NAs:

```
df <- df %>%
        mutate_at(.vars = c(95:104), funs(ifelse(. %in% c('NA', '[NA]'), NA, .)))
```

These columns are supposidely having information coded in a varying number of values, repectively 5, undetermined number, 2, 5, 2, 2, 2, 2, undertermined number, 2. In reality, the columns contain the following number of values:

	Expected number of values	Actual number of values
Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)	5	32
Land use detail (e.g. number of cuts or years since abandonment)	NA	917
Mowing (1/0)	2	3
Mowing frequency: cuts per year (2=2cut/yr; 1=1cut/yr; 0.5=1 cut/2 yr or 2 yr abandoned; 0.2=1 cut/5 yr or 5 yr abandones; 0=never mown)	5	9
Grazing (1/0)	2	5
Grazing intensity (0/1) 1= intensive grazing	2	14
Burning (1/0)	2	3
Ex arable (1/0)	2	2
Ex arable years (years since last ploughing)	NA	5
Fertilized (1/0)	2	4

Details on the land-use information available

Land use (column 95 / column CQ)

Problems

This column contains 32 values instead of 5:

```
## $`Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)`
##  [1] "grazed"                    "grazed, mown"
##  [3] "mown"                      "natural grassland"
##  [5] "abandoned"                 NA
##  [7] "grazed/natural"            "grazed/mown"
##  [9] "grazing"                   "Abandoned"
## [11] "Mown and grazed"           "Mown"
## [13] "mown and grazed"           "Natural grassland"
## [15] "mown/abandoned"            "grazed/abandoned"
## [17] "mown/grazed"               "abandoned?"
## [19] "LAS"                       "P"
## [21] "L"                         "N"
## [23] "grazedazed"                "grazed, burnt"
## [25] "burnt, abandoned"          "abandonment"
## [27] "abondonment"               "0 intensive grazing"
## [29] "0t visible"                "natural grassland, grazed"
## [31] "grazed and mown"           "trampled"
## [33] "irregularly mown"
```

- Some of the values are just typos (*e.g.* capital letter in the beginning, misspelling, extra verbose...).
- Some are actually two or more LU classes combined (*e.g.* grazed/natural...)
- Some are LU classes not listed (*e.g.* trampled, grazed/burnt)
- Some seems to indicate the absence of LU (*e.g.* 0t visible)
- Some are not comprehensible (*e.g.* LAS, P, L, N)

Suggestions

- Fixing the typos
- Deciding whether it is better to have mixed classes, or 5 columns coded as binary variables. If the former is better, then one should homogeneize the way the mixed classed are coded. I suggest to stick to the original order, with different LU separated by a '/'. It is noteworthy that these binary columns already somehow exists (*i.e* there is already Grazing, Mowing, Ex-arable (= abandoned?) binary columns). See below for more details on these columns.
- Either adding classes to matches the new LU classes, matching with the closest class, or removing completely, providing there is no more info in Column 96 (LU details).
- Replacing by NA, are there is no LU visible

Land use details (column 96 / column CR)

Problems

This columns contain 917 unique values, with various information, freely written by data provider. The information is not usable as such, as it is not standardized, for instance:

```
## [1] "very extensive grazing"
## [2] NA
## [3] "mowing once a year at the end of July"
## [4] "Extensive grazing by goats"
## [5] "Low intensity horse pasture, montane meadow"
## [6] "Low intensity pasture (cattle, horses)"
```

Suggestions

Manually create a lookup table to match this columns and the other. One should try to extract intensity information when possible.

Mowing 0/1 (column 97 / column CS)

Problems

- Some plots have impossible values ('?')
- Some plots have mown = 1 but are classified as abandoned or natural grassland in column 95
- Some plots with mown = 0 are classified as mown in column 95:
- Some plots with mown = NA are classified as mown in column 95

Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)	?	0	1	<NA>
0 intensive grazing	NA	NA	NA	1
0t visible	NA	NA	NA	2
abandoned	NA	858	12	15194
Abandoned	NA	NA	NA	67
abandoned?	NA	6	NA	NA
abandonment	NA	NA	NA	1
abondonment	NA	NA	NA	4
burnt, abandoned	NA	13	NA	NA
grazed	NA	3794	NA	30153
grazed and mown	NA	NA	4	NA
grazed, burnt	NA	NA	NA	26
grazed, mown	NA	NA	128	12
grazed/abandoned	NA	NA	NA	23
grazed/mown	NA	NA	NA	15
grazed/natural	NA	NA	NA	64
grazedazed	NA	NA	NA	1
grazing	NA	NA	NA	10
irregularly mown	NA	NA	1	NA
L	NA	NA	NA	12
LAS	NA	NA	NA	12
mown	NA	4	430	6578
Mown	NA	NA	34	2
mown and grazed	NA	NA	NA	178
Mown and grazed	NA	NA	NA	15
mown/abandoned	NA	NA	NA	10
mown/grazed	NA	NA	7	NA
N	NA	NA	NA	28
natural grassland	NA	1562	4	104572
Natural grassland	NA	17	NA	NA
natural grassland, grazed	NA	NA	NA	78
P	NA	NA	NA	48
trampled	NA	1	NA	NA
NA	6	22	NA	16068

Suggestions

- Deciding which column takes precedence (column 95 or binary column), and reclassify accordingly. To do so, plots (or datasets) with discrepancies need to be identified and the original tables need to be manually checked.
- Plots that are not mown should be identified with mown = 0 (*i.e* not NAs in this column)

Mowing frequency (column 98 / column CT)

Problems

Some values (4, 0.3, 0.03, 0.05, 0.1) do not match the values that should be present. In addition, some plots listed a mown = 0 have a mowing frequency:

Mowing frequency: cuts per year (2=2cut/yr; 1=1cut/yr; 0.5=1 cut/2 yr or 2 yr abandoned; 0.2=1 cut/5 yr or 5 yr abandones; 0=never mown)	?	0	1	<NA>
0	NA	468	NA	NA
0.03	NA	16	NA	NA
0.05	NA	159	NA	NA
0.1	NA	30	NA	NA
0.2	NA	NA	4	NA
0.3	NA	4	8	NA
0.5	NA	NA	31	NA
1	NA	4	212	NA
4	NA	NA	15	NA
NA	6	5596	350	173174

Suggestions

- Checking in the column 96 if the frequency information can be corrected
- Reclassifying plots with a mowing freaquency as mown = 1 (*i.e.* the frequency information takes precedences over the binary info. This rule should stand for all the frequency information)
- If no more info can be found, checking with Idoia

Grazing (1/0) (column 99 / column CU)

Problems

- Some plots have impossible values ('?', 'probably' or '2')
- Some plots with grazing = 1 are listed as abandoned, mown or natural grassland
- Some plots with grazing = NA are listed as grazed or mixed grazed

Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)	?	0	1	2	probably	<NA>
0 intensive grazing	NA	NA	NA	NA	NA	1
0t visible	NA	NA	NA	NA	NA	2
abandoned	NA	1645	38	NA	NA	14381
Abandoned	NA	NA	NA	NA	NA	67
abandoned?	NA	6	NA	NA	NA	NA
abandonment	NA	NA	NA	NA	NA	1
abondonment	NA	NA	NA	NA	NA	4
burnt, abandoned	NA	NA	NA	NA	NA	13
grazed	NA	NA	5103	1	NA	28843
grazed and mown	NA	NA	NA	NA	NA	4
grazed, burnt	NA	NA	26	NA	NA	NA
grazed, mown	NA	NA	130	NA	NA	10
grazed/abandoned	NA	NA	NA	NA	NA	23
grazed/mown	NA	NA	NA	NA	NA	15
grazed/natural	NA	NA	NA	NA	NA	64
grazedazed	NA	NA	NA	NA	NA	1
grazing	NA	NA	NA	NA	NA	10
irregularly mown	NA	NA	NA	NA	NA	1
L	NA	NA	NA	NA	NA	12
LAS	NA	NA	NA	NA	NA	12
mown	NA	176	39	NA	6	6791
Mown	NA	NA	34	NA	NA	2
mown and grazed	NA	NA	NA	NA	NA	178
Mown and grazed	NA	NA	NA	NA	NA	15
mown/abandoned	NA	NA	NA	NA	NA	10
mown/grazed	7	NA	NA	NA	NA	NA
N	NA	28	NA	NA	NA	NA
natural grassland	NA	1349	60	NA	NA	104729
Natural grassland	NA	17	NA	NA	NA	NA
natural grassland, grazed	NA	NA	NA	NA	NA	78
P	NA	NA	NA	NA	NA	48
trampled	NA	NA	NA	NA	NA	1
NA	6	17	NA	NA	NA	16073

Suggestions

- Decide which column takes precedence (column 95 or binary column), and reclassify accordingly. I have no strong opinion on that, we should check which information was provided first
- Plots that are not grazed should be identified with grazed = 0 (*i.e* not NAs in this column)

Grazing intensity

Problems

- We have no information on the values that should be in this column
- Some information are coded as plain text
- Some plots with grazing = 1 have no grazing frequency:

Grazing intensity (0/1) 1= intensive grazing	?	0	1	2	probably	<NA>
0	NA	1611	219	NA	NA	NA
0.1	NA	NA	202	NA	NA	NA
0.25	NA	NA	448	NA	NA	NA
0.3	NA	NA	92	NA	NA	NA
0.5	NA	NA	749	NA	NA	8
0.75	NA	NA	148	NA	NA	NA
1	NA	NA	2961	NA	NA	NA
2	NA	NA	28	NA	NA	NA
3	NA	NA	31	NA	NA	NA
7	NA	NA	4	NA	NA	NA
high	NA	NA	12	NA	NA	NA
low	NA	NA	15	NA	NA	NA
middle	NA	NA	30	NA	NA	NA
overgrazing	NA	NA	3	NA	NA	NA
NA	13	1627	488	1	6	171381

Suggestions

- Checking what values should be there with Idoia

Burning (1/0) (column 101 / column CW)

Problems

- Plots classified as burnt (grazed, burnt and movn, burnt) have burn = NA

Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)	?	0	1	<NA>
0 intensive grazing	NA	NA	NA	1
0t visible	NA	NA	NA	2
abandoned	NA	658	29	15377
Abandoned	NA	NA	NA	67
abandoned?	6	NA	NA	NA
abandonment	NA	NA	NA	1
abondonment	NA	NA	NA	4
burnt, abandoned	NA	NA	NA	13
grazed	NA	3649	52	30246
grazed and mown	NA	4	NA	NA
grazed, burnt	NA	NA	NA	26
grazed, mown	NA	7	14	119
grazed/abandoned	NA	NA	11	12
grazed/mown	NA	NA	NA	15
grazed/natural	NA	NA	NA	64
grazedazed	NA	NA	NA	1
grazing	NA	NA	NA	10
irregularly mown	NA	1	NA	NA
L	NA	NA	NA	12
LAS	NA	NA	NA	12
mown	NA	361	11	6640
Mown	NA	34	NA	2
mown and grazed	NA	NA	NA	178
Mown and grazed	NA	NA	NA	15
mown/abandoned	NA	NA	NA	10
mown/grazed	NA	7	NA	NA
N	NA	28	NA	NA
natural grassland	NA	1596	17	104525
Natural grassland	NA	17	NA	NA
natural grassland, grazed	NA	NA	NA	78
P	NA	NA	NA	48
trampled	NA	1	NA	NA
NA	NA	62	3	16031

Suggestions

- Reclassify those plots as burn = 1

Ex arable (1/0) (column 102 / column CX)

Problems

- Plots classified as abandonned have ex arable = 0 or ex arable = NA

Land use (5 standard categories: mown, grazed, abandoned, natural grassland, NA)	0	1	<NA>
0 intensive grazing	NA	NA	1
0t visible	NA	NA	2
abandoned	317	17	15730
Abandoned	NA	NA	67
abandoned?	NA	NA	6
abandonment	NA	NA	1
abondonment	NA	NA	4
burnt, abandoned	NA	NA	13
grazed	3081	6	30860
grazed and mown	NA	NA	4
grazed, burnt	NA	NA	26
grazed, mown	NA	20	120
grazed/abandoned	NA	NA	23
grazed/mown	NA	NA	15
grazed/natural	NA	NA	64
grazedazed	NA	NA	1
grazing	NA	NA	10
irregularly mown	1	NA	NA
L	NA	NA	12
LAS	NA	NA	12
mown	153	30	6829
Mown	NA	NA	36
mown and grazed	NA	NA	178
Mown and grazed	NA	NA	15
mown/abandoned	NA	NA	10
mown/grazed	NA	NA	7
N	NA	NA	28
natural grassland	1354	13	104771
Natural grassland	NA	NA	17
natural grassland, grazed	NA	NA	78
P	NA	NA	48
trampled	1	NA	NA
NA	37	11	16048

Suggestions

- Checking the definition of abandonement and ex arable used in the database

Ex arable years (column 103 / column CY) and Fertilized (1/0) (column 104 / column CZ)

This column should be checked against the column 96 manually to ensure the validity of the data. Note that there are impossible values for the Fertilized column (0.3 and 0.5) If the data were provided as is by the data owner, there is not much we can do (the data owner could have provided that information without filling the information in column 96)