General description of the dataset

The experimental design of the QUF Project consists of 64 trees of the species *Pinus Pinea* and *Quercus Ilex* distributed in 4 control plots where 4 treatments are applied (for further details see the "White book"):

1. Absence of treatment or witness.
2. Retainer.
3. Mycorrhiza.
4. Mixed treatment (retainer + mycorrhiza).

The observation period runs from September 2014 to November 2016. During this period moisture and temperature data in 2 depths (20 and 40 cm) of those 64 trees has been collected every half hour. Additionally, weather data of the whole plantation has been collected every half hour.

The number of observations accumulated up to November 30, 2016 is 2.525.312.

Description of the data

Within the dataset there are 3 main groups of variables: identification variables (measurement period, identification of the tree, block, treatment applied and specie of the tree); measurement variables (humidity and temperature, both at 20 cm and at 40 cm depth); and, finally, climate variables (environment humidity and temperature, rainfall and wind speed). The variables are as follows:

1. Identification variables:
    - Period: date and time of the measures taken. The data is collected every half hour from September 1, 2014 until November 30, 2016. Therefore, there are 48 observations per tree per day.
    - Tree ID: identification of each tree on which measurements are taken. There are 64 trees.
    - Block: control plot where the tree is planted. There are 4 control plots in our experimental design (1, 2, 3 and 4).
    - Specie: specie of the tree. 2 species: *Pinus Pinea* (P) and *Quercus Ilex* (Q).
    - Treatment: each of the 4 different treatments applied to the trees. Treatment with mycorrhiza (Mic), retainer (Ret), mixed (Mix), or no treatment (No).

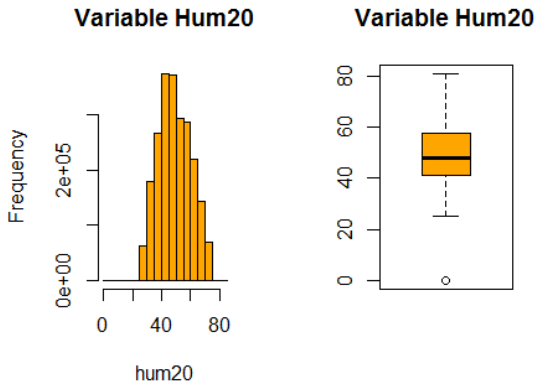
2. Measurement variables:
    - Hum20: indirect measure of percentage of soil moisture at 20 cm. depth in the tree (see White book for details).
    - Hum40: indirect measure of percentage of soil moisture at 40 cm. depth in the tree (see White book for details).
    - Tem20: temperature measured in ºC to 20 cm. of depth in the tree.
    - Tem40: temperature measured in ºC to 40 cm. of depth in the tree.

3. Climate variables:
    - Humidity: percentage of air moisture measured in the plantation.
    - Temperature: environment temperature measured in ºC in the plantation.
    - Precipitation: amount of rain measured in mm in the plantation in the last half hour.
    - Wind speed: wind speed measured in km/h in the plantation.

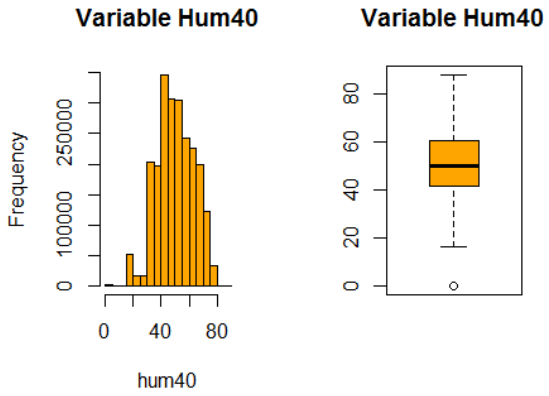
Quantitative description of the data

Both the measurement variables and the climatological variables are described:

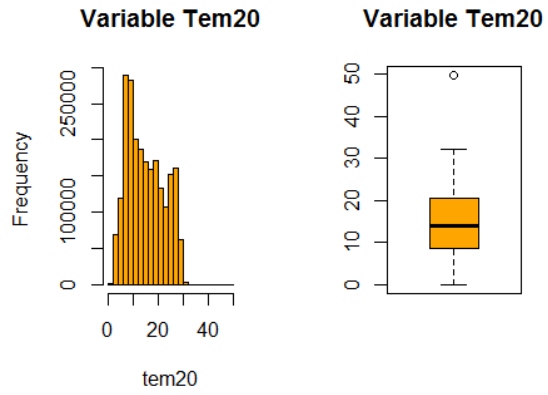
1. Measurement variables:
   - Hum20: the average indirect moisture percentage at 20 cm depth is 49,17. There are 255.745 missing values (NA's) out of the 2.525.312 total measures. The histogram and boxplot of the Hum20 is as follows:



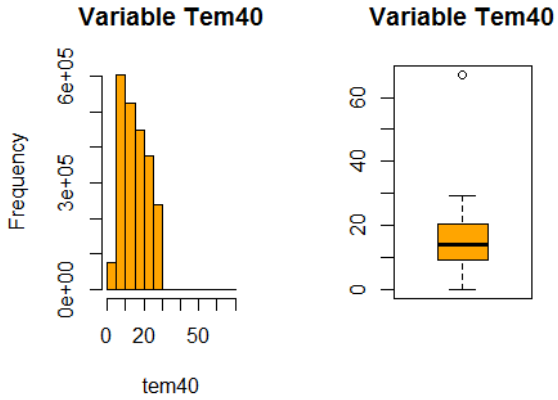
   - Hum40: the average indirect moisture percentage at 40 cm depth is 50,23. There are 255.744 missing values (NA's) out of the 2.525.312 total measures. The histogram and boxplot of the Hum40 is as follows:



   - Tem20: the average soil temperature at 20 cm depth is 14,98 °C. There are 255.801 missing values (NA's) out of the 2.525.312 total measures. The histogram and boxplot of the Hum20 is as follows:



- Tem40: the average soil temperature at 40 cm depth is 15,09 ºC. There are 255.801 missing values (NA's) out of the 2.525.312 total measures. The histogram and boxplot of the Hum40 is as follows:



2. Climate variables:
    - Humidity: the average environment humidity percentage in the plantation is 68,11 %. There are 51.648 missing values (NA's) out of the 2.525.312 total measures.
    - Temperature: the average environment temperature in the plantation is 12,56 ºC. There are 51.648 missing values (NA's) out of the 2.525.312 total measures.
    - Precipitation: The average rainfall per half hour in the plantation is 0,03 mm.
    - Wind speed: The average wind speed in the plantation is 19,40 km/h.

Limitations

To our best knowledge, these are the most important limitations of the dataset:

1. When analyzing the data, it is important to carefully consider potential outliers. For instance, the humidity at 40cm. of the tree "01-13-P-1" it is likely to be problematic because of its lower values and smaller variation compared to the other trees. We have also found problems in the measures of the wind speed during the months of May and June 2016. It is likely that under specific conditions the probes are collecting mistaken measurements.
2. There are several missing values for specific trees and for specific dates. For instance, during August 2015 there was a problem with the communication network and several observations were missed.
3. We cannot assure that the date and time of the observations is accurate. We have observed a problem at the end of the project with the data and time collected in the data being different that the true date and time of the observation. Therefore, if the findings of the analysis depend heavily on the date and time the results should be taken carefully.