DATASET AND DATA PROCESSING DETAILS

THE LONG ROAD TO SOBRIETY: ESTIMATING THE OPERATIONAL POWER CONSUMPTION OF CELLULAR BASE STATIONS IN FRANCE

Arsalan Ahmed and *Marceau Coupechoux*

Information Processing and Communications Laboratory (LTCI)
Télécom Paris

October 29, 2022

1 ANFR DATASET

1.1 Description

The utilized public dataset on radioelectric installations is from the ANFR which is a governmental agency managing all radio frequencies in France. This dataset is developed and gathered by the ANFR as they are tasked with the authorization of all transmission site deployments (above 5 Watts) in mainland France as well as in overseas French territories. This publicly accessible dataset provides a unique opportunity for researchers to conduct different studies on national scale.

The following are the key points regarding the dataset:

- The dataset includes data from all public and private operators of radio installations.
- The dataset contains data from April 2015 and is updated on a monthly basis.
- Every month two folders are uploaded to the dataset, one for the actual data files while the other for the reference files.
- The files contained within each of the monthly folders are shown in Table 1.

Table 1: Files within each monthly dataset folders

Data Files	Reference Files
SUP_ANTENNE.txt	SUP_EXPLOITANT.txt
SUP_BANDE.txt	SUP_NATURE.txt
SUP_EMETTEUR.txt	SUP_PROPRIETAIRE.txt
SUP_STATION.txt	SUP_TYPE_ANTENNE.txt
SUP_SUPPORT.txt	

The contents within each of the files and the relationship among them are explained next.

1.2 Contents:

It is to be noted that the details mentioned in this section are verified through correspondence with the ANFR itself. The contents within each of the files are indicated by the filenames. Each file is a semicolon separated text file containing tabular data. The information contained within each of the files along with the associated reference files is explained below:

- SUP_SUPPORT.txt: A support is a physical location/site usually a building, tower or pylon on which antennas are installed. Hence, the SUP_SUPPORT.txt file contains information on all the existing supports. The columns in this file are described in Table 2.

Table 2: Columns in SUP_SUPPORT.txt

Column Name	Description
SUP_ID	Unique ID for support identification.
STA_NM_ANFR	ID for identifying the stations installed on the support. This field allows the join operation with the SUP_STATION table.
NAT_ID	ID for recognizing the type/nature of the support. This field is numbered and the corresponding labels can be found in the reference SUP_NATURE.
COR_NB_DG_LAT COR_NB_MN_LAT COR_NB_SC_LAT COR_CD_NS_LAT COR_NB_DG_LON COR_NB_MN_LON COR_NB_SC_LON COR_CD_EW_LON	Geographical coordinates of the support in the World Geodetic System (WGS) 84.
SUP_NM_HAUT	Height of the support.
TPO_ID	ID for recognizing the owner of the support. This field is numbered and the corresponding labels can be found in the reference SUP_PROPRIETAIRE.
ADR_LB_LIEU ADR_LB_ADD1 ADR_LB_ADD2 ADR_LB_ADD3	Location address of the support.
ADR_NM_CP	Postal code of the support.
COM_CD_INSEE	INSEE code of the support.

- SUP_STATION.txt: A station or Base Station (BS) is a set of transceivers owned by a single operator. Hence, the SUP_STATION.txt file contains information on all the existing BSs. The columns in this file are described in Table 3.

Table 3: Columns in SUP_STATION.txt

Column Name	Description
STA_NM_ANFR	Unique ID for BS identification. This field allows the join operation with all the other data tables.
ADM_ID	ID for recognizing the operator of the BS. This field is numbered and the corresponding labels can be found in the reference SUP_EXPLOITANT.
DEM_NM_COMSIS	ID for identifying the Cartoradio installation. Cartoradio is a cartographic tool provided by the ANFR. Cartoradio allows to view all the radio sites in France on an interactive map which displays their location along with measurements of the exposure to their electromagnetic waves [1].
DTE_IMPLANTATION	Date on which the BS is authorized by the ANFR.
DTE_MODIF	Date on which the most recent technical modification is approved by the ANFR.
DTE_EN_SERVICE	Date on which the BS became *in service*. This is based on the commissioning date of the first transceiver. Usually, a BS is first authorized by the ANFR and then, later put *in service* when the first transceiver is commissioned.

- SUP_ANTENNE.txt: An antenna is a metallic structure that can capture as well as transmit radio electromagnetic waves. Thus, it is a vital component of a BS transceiver providing the users with cellular connectivity. Hence, the SUP_ANTENNE.txt file contains information on all the existing antennas. The columns in this file are described in Table 4.
- SUP_EMETTEUR.txt: The actual translation for the word EMETTEUR is transmitter. However, it is to be pointed out that here EMETTEUR is actually a *transceiver*. A transceiver consists of a transmitter and receiver inside a single unit. Hence, the SUP_EMETTEUR.txt file contains information on all the existing transceivers. The columns in this file are described in Table 5.

Table 4: Columns in SUP_ANTENNE.txt

Column Name	Description
STA_NM_ANFR	ID for identifying the BS to which the antenna belongs. This field allows the join operation with the SUP_STATION table.
AER_ID	Unique ID for antenna identification. This field allows the join operation with the SUP_EMETTEUR table.
TAE_ID	ID for recognizing the type of the antenna. This field is numbered and the corresponding labels can be found in the reference SUP_TYPE_ANTENNE.
AER_NB_DIMENSION	Size of the antenna.
AER_FG_RAYON	Type of radiation (D for directional and N for omnidirectional).
AER_NB_AZIMUT	Azimuth of the directional radiation.
AER_NB_ALT_BAS	Antenna height relative to the ground.
SUP_ID	ID for identifying the support on which the antenna is mounted on.

Table 5: Columns in SUP_EMETTEUR.txt

Column Name	Description
EMR_ID	Unique ID for transceiver identification. This field allows the join operation with the SUP_BANDE table.
EMR_LB_SYSTEME	Label for the wireless communications standard/technology.
STA_NM_ANFR	ID for identifying the BS to which the transceiver belongs to. This field allows the join operation with the SUP_STATION table.
AER_ID	ID for identifying the antenna which the transceiver uses. This field allows the join operation with the SUP_ANTENNE table.
EMR_DT_SERVICE	Commissioning date of the transceiver.

- SUP_BANDE.txt: A frequency band is a specific range of frequencies in the RF spectrum over which transmissions are conducted. Each transceiver operates at specific frequency bands whose information is contained in the SUP_BANDE.txt file. It is to be noted that a single-band Frequency Division Duplex (FDD) transceiver is associated with two bands in the SUP_BANDE table; one for the uplink while the other is for the downlink. For Time Division Duplex (TDD) there is only one band which is used for both transmission and reception. Also, there are multi-band transceivers associated with multiple frequency bands. The columns in this file are described in Table 6.

Table 6: Columns in `SUP_BANDE.txt`

Column Name	Description
`STA_NM_ANFR`	ID for identifying the BS to which the frequency band belongs to. This field allows the join operation with the `SUP_STATION` table.
`BAN_ID`	Unique ID for frequency band identification.
`EMR_ID`	ID for identifying the transceiver which uses this band. This field allows the join operation with the `SUP_EMETTEUR` table.
`BAN_NB_F_DEB`	Starting frequency of the band.
`BAN_NB_F_FIN`	Ending frequency of the band.
`BAN_FG_UNITE`	Unit of the mentioned frequencies (K for KHz, M for MHz and G for GHz).

1.3 Hierarchy:

There exists a hierarchy in the dataset which should be used to combine the individual data tables into one large table providing an overall view of the information. The hierarchy is depicted in Figure 1.

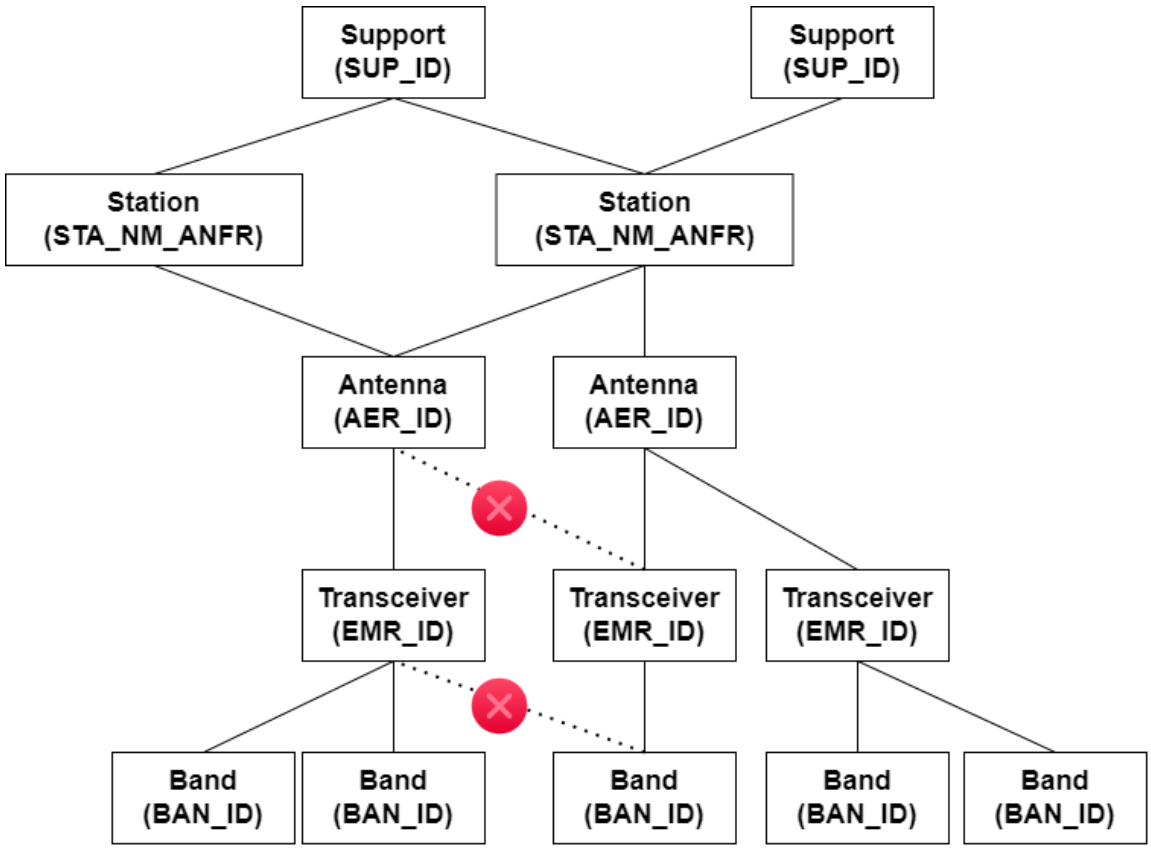

Figure 1: Hierarchy of the ANFR dataset

From Figure 1, it can be seen that a support (also known as a site) is at the top of the hierarchy. Each support/site consists of several stations which can belong to the same or different operators. In some rare cases, a single BS can be attached to multiple support/sites. For example, if a BS uses two antennas installed on a large building and if these antennas are more than 90 meters apart then the BS is said to be supported by two different supports. However, for some cases like a stadium spanning more than 90 meters, the establishments are not divided into several supports.

A BS is owned by a single operator and uses one or more antennas. It is also possible that several BSs use the same antenna. Sharing of telecom infrastructure is becoming quite popular where competitors become partners to lower their costs. This trend is present in France as well where leading mobile operators have signed mutual strategic agreements to share their resources among each other. Specifically, an agreement between Bouygues Telecom and SFR from 2014 allows them to maintain shared BS assets [2]. From the dataset, this seems a case of passive infrastructure sharing where non-electronic equipment at a cell site is shared. An antenna is associated with one or more transceivers. However, one transceiver can only use a single antenna.

A transceiver runs only a single type of wireless communications standard/technology, for example, `GSM 900`, `LTE 1800`, `5G NR 700` etc. Every transceiver is associated to one or more frequency bands but these bands can be from the same wireless standard only.

2 DATASET MANIPULATION

2.1 Coding Structure:

The programming language that is used for dataset processing is python. Python's large collection of libraries and packages allow for simple and faster code development. Specifically, python's powerful pandas library is mostly utilized because of its support for cleaning, manipulating and analyzing tabular data. The pandas 2D data structure of a Dataframe similar to a 2D array or a table with rows and columns is particularly useful. The two stages of processing the dataset includes:

1. Dataset Preprocessing
2. Data Analysis

2.2 Data Preprocessing:

The dataset size is around 17 GB (till September 2022) which is quite large. However, not all information contained within the dataset is useful and relevant to the power consumption of cellular BSs in France. Hence, the goal of this step is to preprocess the dataset, filter out the relevant information only and reduce the dataset size. The compact preprocessed dataset can then be utilized for faster data analysis and power consumption related calculations.

The following are the assumptions and considerations taken during the preprocessing stage:

1. The `SUP_SUPPORT.txt` file is not used in the preprocessing code since it only contains information regarding the physical support and location of the BSs which is irrelevant for power consumption calculations.
2. The rest of the tables are joined together based on the mentioned hierarchy. Furthermore, using the mapping contained within the reference tables, the associated columns are replaced within the main (actual data) tables.
3. As the goal is to estimate the national power consumption of cellular BSs in France, hence only the cellular technologies/standards are considered. These include `GSM 900`, `GSM 1800`, `GSM 900/1800`, `UMTS 900`, `UMTS 2100`, `UMTS 2100/900`, `LTE 700`, `LTE 800`, `LTE 1800`, `LTE 2100`, `LTE 2600`, `NR 700`, `NR 2100` and `NR 3500`. Overall, all the mobile communication generations are considered but their experimental bands like `LTE 1400 Expe` etc. are filtered out.

4. Out of all the telecom operators in France, only the four main operators are considered which include **Orange, SFR, Bouygues Telecom** and **Free Mobile**.

5. Only the *in service* BSs are considered. This is determined from the `DTE_EN_SERVICE` column in the `SUP_STATION` table. If this field is empty, it means that the BS is authorized by the ANFR but not *in service* yet. Hence, only the rows with a valid `DTE_EN_SERVICE` are used when analyzing the data.

A single preprocessed text file is generated for each month. The columns in the preprocessed file are described in Table 7.

Table 7: Columns in the preprocessed file

Column Name	Description
`BS_ID`	ID for identifying the BS.
`DoS`	Date on which the BS became *in service*.
`OPR_NAME`	Name of the telecom operator.
`ANT_TYPE`	Type of the antenna used by the transceiver.
`TRX_ID`	Unique ID per line to identify the transceiver.
`System`	Label for the wireless communications standard/technology.
`FDD_BW`	Aggregate FDD bandwidth of the transceiver. The mentioned bandwidth is half of the total FDD bandwidth.
`TDD_BW`	Aggregate TDD bandwidth of the transceiver. The mentioned bandwidth is the total TDD bandwidth.
`BW_UNIT`	Unit of the mentioned bandwidth.

Every line in a preprocessed text file corresponds to one unique transceiver. This compact and condensed information is quite vital from the perspective power consumption as every transceiver is subjected to a different model depending on their cellular system, bandwidth and duplexing mode.

REFERENCES

[1] ANFR, "Cartoradio: The map of radio sites and wave measurements." https://www.cartoradio.fr/#/. Accessed: August 2022.

[2] Bouygues Telecom, "Bouygues Telecom and SFR conclude a strategic agreement to share a part of their mobile access networks." https://www.bouygues.com/wp-content/uploads/2014/01/14_0019_20140131_CP_ByT_SFR_Mutualisation-vd%c3%a9f_va.pdf, 2014. Accessed: August 2022.