Details regarding the creation of loads-related input tables for SWITCH Mexico Model

Aldo Sayeg Pasos Trejo[1,2]

[1]*Physics Departament. Facultad de Ciencias. Universidad Nacional Autonoma de Mexico*
[2]*Visiting Student Researcher for the Berkeley Energy and Climate Institute at University of California, Berkeley*

August 20, 2016

1 Creating the 'Hourly Demand' table and associated problems

Inside the "Loads/Data Analysis" directory of the "Switch_mexico_data" repository, there is a directory for each of the hourly load projections in Mexico. There are high, mid and low projections. Each one of these projections is made by CENACE ('Centro Nacional de Control de Energia', National Center for energy control)[**?**].

These projections consist of a table with the hourly demand for every Control Region of the SEN ("Sistema energetico nacional", national energy system) from 2016 to 2030. On the other hand, SENER ("Secretaría de energía", Department of energy) has also a table of the percentual assigment of load for each load zone within every Control Region of the SEN. With this information, we were able to make a table consisting of the hourly demand for all the SEN load zones from 2016 to 2030 in each of the 3 projections.

For an easier manipulation of the code, the scripts and tables made for all three projections have the same name and are diferentiated from each other by being placed on separate directories. Each projection directory has the name "Net load per hour in * scenario", where * can be high, mid or low. We will continue to use the sign "*" throughout this report when we are referring to any of these three words that specify the projection scenario. The switch-nicaragua and switch input tab files created are in other directories, so those files do have a diferent name depending on the projection they were based on.

Due to the amount of data, colecting all of this in a single cvs file would have been too large for efficient manipulation inside Python. Also, as these tables were made manually due to the fact that the original format was an xlsx file with several sheets, it would have been really long to create a single file with all the data at this point.

As a result of these challenges, we split this "hourly demand per load zone" table in 15 csv files contained in the "RawTables" folder of each projection directory, one for each year from 2016 to 2030. Each of the files has the name of the year of which it contains data from. All of these tables were made manually in Excel, and then saved to the desired format.

However, we still had problems with these manually made tables. These problems were related to data parsing, such as missing information and changes in the load zone number and name:

- The information regarding the "31-central" load zone was missing from the raw tables created manually.

- A load zone in the system, called "20-tamazunchale" was not included in these published files for some unexplained reason. We decided to include this load zone and assigned it a percentage participation factor for its control region.

- Confusion between the difference between a pair of load zones called "53-mulege" and "54-loreto". These load zones are marked as the same in some goverment documents and as different in some other.

Due to this, we used a Python script called "Organized tables creation" to read the tables contained in the "RawTables" directory and then make the corrections regarding these aforementioned problems.

2 Python scripts

2.1 Problem solutions and "Organized table creation"

The "Organized table creation" Python script of each projection directory creates the final and corrected hourly-load tables, using as a base the manually created tables in the "RawTables" folder. First of all, it reads each one of the tables contained in the "RawTables" folder. The script appends each table, along its corresponding year, into a single Data Frame. After that, it reads the information corresponding to the "31-central" load zone from the "RawTables/Central.csv" file and attaches it to the Data Frame.

To fix the "20-tamazunchale" missing load zone problem, we had to assign it an hourly load value. We decided to take approximately 2% of the total load of its balancing area ("06-noreste") and assign it to it. This 2% was taken from the "16-monterrey" load zone, due to the fact that this "16-monterrey" load zone participates with a little more than 50% of the load in the "06-noreste" balancing area.

After comunication with Mexico's SENER, they corrected the problem regarding the "53-mulege" and the "54-loreto load zones". They determined that the load zone "54-loreto" doesn't exist and that its participation factor within its balancing area is distributed proportionally between the other load zones at that balancing area ("09-baja_california_sur-la_paz")

Finally, after making the adjusments to fix these problems, the script exports a new .csv table for each year of the hourly-load. These tables are located inside the "OrganizedTables" folder of each projection directory.

2.2 Other scripts

Also we made a Python script called "Statistical analysis table creation" that creates some desired tables. First of all, the script creates a single data frame called "df" that contains the data of every year (a concatenation of the files in the "OrganizedTables" folder). It exports this file to the "OranizedTables" folder as a large csv file called "HourlyLoadPerNode".

The same script creates another table, called "LoadHighlightsPerNode" that calculates, for every load zone at each month and year from 2016 to 2030, the peak day of the month ("PeakDay" column), the peak day average load ("PeakDayValue" column) and the average of the month ("MonthlyAverage" column). All of these columns are in Megawatts. We will procede to explain how were these columns calculated.

2.2.1 "LoadHighlightsPerNode" data details

For every load zone in the columns of the "LoadHighlightsPerNode" and a given row of year and month, the "PeakDay" column corresponds to the day of that month and year that has the highest hourly load value for a given hour. The "PeakDayValue" shows that maximum load value.

On the other hand, the "MonthlyAverage" value calculates the daily average of the daily load for every day in that month and year, but excluding the peak day. For example, if at load zone "Hermosillo" at year 2016 and month 09, the peak day is 03, the Monthly average column will correspond to the sum of the daily load of days 01,02,04,05,...,n (n is the number of days in month 09 at year 2016) divided by n-1 (as we have excluded the peak day, we hace only summed n-1 days, so for the average of these sum we have to divide by n-1)

The Script called "Graphical Analysis" displays graphics corresponding to this data. The first part displays the average hourly load for every day in a year, for each year from 2016 to 2030, in a separate graphic ("figure", in matplotlib's terms) for every load zone in the SEN.

The second part asks you for an input consisting of a load zone, a Month and a year. For that informations, it outputs two graphics: one consisting of the hourly load for every day in the month on that node and another one of the hourly load for selected days.

These selected days are the following: the "PeakDay" from the "LoadHighlightsPerload zone" table, the hourly average median day of the month (the day of the month whose hourly average load is the median compared to all the other days hourly average load), the day whose hourly average is closest to the "MonthlyAverage" value from "LoadHighlightsPerNode" table, and a random day

The script called "Switch-nicaragua input creation" creates a table called "la_hourly_demand_*" . This table is saved in the "OrganizedTables" folder of each projection directory and also at the "Main Tabs" directory of the "switch_mexico_data" repository.

We made this table with the same format as the table contained in the SWITCH Nicaragua input directory. However, it is not the final SWITCH input table, as we later in the project realized. For that reason, we created the "switch loads creation" table, from who we will talk in the next subsection.

2.3   "switch loads creation" script

The "switch loads creation" python script creates tables that the SWITCH model uses as an input. These SWITCH input tables created resemble the tables called "loads.tab" and "lz_peak_loads.tab" located in the "switch_mod/test_dat" directory of the switch-master repository[?].

The "loads.tab" table consists of a list of the hourly load in each of the load zones in an specific format of date. This date format consists of a single string of numbers where the digits represent the year, month, day and hour (hour in 24 hour format) of that load in that load zone. The date in this format is contained in the "timepoint" column. Instead of having different columns for each load zone load, the hourly loads are represented in a single column called "lz_demand_mw" and the table uses a column called "load_zones" specifing the load zone to which that hourly load belongs.

At the end, the "loads_*.tab" files consist of a table of 3 columns and nearly 6,000,000 rows. Because of this, the file size is too large to be hosted on the "switch_mexico_data" GitHub repository.

The "lz_peak_zones_*.tab" files consists of a table indexed by the load zones and two columns. The "peak_demand_mw" column corresponds to the hourly peak demand on that load zone for the period of time that the "period" column specifies. The "period" columns shows a code for the period of time from wich that hourly-load peak value is taken. The period of time is coded as in the following example:

"period" column name	period of time that represents
2016	2016-2020
2021	2021-2025
2026	2026-2030

2.4   Loads .tab files created outside of any script.

The "load_zones.tab" file located in the consists of three columns indexed by the load zone. The "cost_multipliers" column refers to a number that is multiplied to the estimated distribution to cost to compensate for losses due to several factors, such as damage of the transmission or distrbution network and the electricity thefts from clandestine networks. The "existing_local_td" corresponds to the maximum demand for that load area in megawatts during the 2016-2020 period. The last column, named "local_td_annual_cost_per_mw" corresponds to an estimation of the distribution costs in that load area in a year, in USD, normalized by the megawatts generated by the load zone in that year. Further details on how the estimation of distribution costs were made are contained in the technical report of the distribution cost estimation.

References

[1] http://www.gob.mx/sener/acciones-y-programas/programa-de-desarrollo-del-sistema-electrico-nacional-33462? idiom=es

[2] SWITCH master repository https://github.com/switch-model/switch