# input data

Newave is one of the models used by the Electrical System National Operator (ONS) to operate the Brazilian electrical system. Its objective function is to minimize the total system costs considering constrains on load balance, transmission line capacities, operational and investment conditions and other factors.

Every month ONS publishes the deck of data used in the respective run of Monthly Planning Operation (PMO). It contains information about the hydro and thermal power plants, concerning capacity and availability factors, reservoir limits, location of each plant, information about the existent and non-existent plants that are planned to start operation during the period of the run, among others.

COPA reads a file named investOpts.csv. It contains thermal power plants information. In order to create this file, it is necessary the reading of the following files of Newave’s deck of May 2017:

* Term.dat: nominal capacity, capacity factor, availability and minimal generation;
* Clast.dat: variable costs and fuel of thermal power plants;
* Conft.dat: names, localization and situation of each thermal power plant.

Five scripts were created for the cited purposes. They were made in R with the packages “dplyr”, “tidyverse”, “tibble” and “ggplot2”. In a few words, they read the data, clean and organize it in intermediate tables. Finally, there is a script to create the investOpts.csv file for the COPA Brazilian version. The following pages contain a quick description of each script and the main decisions that I had to make during the process.

# scripts

## Reading scripts

1. Reading\_term\_newave.R: It reads nominal capacity, capacity factor, availability and minimal generation from term.dat file.
2. Reading\_clast\_newave.R: It reads unitary variable cost, plants’ situation and fuel of each power plant from clast.dat file, removing redundant information.
3. Reading\_conft\_newave.R: It reads the names, subsystems and status of each thermal power plant from conft.dat file. The status can be NE (non-existent), EX (existing) or EE (existing with future expansion).
4. Reading\_exph\_Newave\_2012\_hydro\_pp\_removed.R: It reads “exp.dat” and detects hydro plants that starts operation after December 2012. The result is a table containing code, names and power of each plants that will enter after 2012. The table is “exph12.csv” and it is saved on “Google Drive @PPE\!IIASA\COPA Initial Data\Hydro”.

All of these scrips create some auxiliary tables to be used in the construction of investOpts.csv file for the Brazilian case.

## Investment\_cost\_thermal.R

It aggregates the information of previous scripts. Additionally, it verifies which plants do not exist yet, and it adds investment cost to them. The results consist in some tables with the required information to be input in COPA, but not yet in the right format.

## Creating\_invest\_opts\_br.R

It reads the tables that was resulted from the previous script (Investiment\_cost\_thermal.R) and creates the file investOpts.csv, in the right format to be used by COPA model.

## Creating\_load\_br

It reads hourly Swedish and Brazilian load information and manipulates them in order to achieve COPA format reading. The period of Brazilian load is 01/01/2015 until 01/01/2017. The result is saved as “load\_Br.feather” and “load\_Br.csv”.

## Creating\_lineCapacities\_br

This script creates the lineCapacities with brazilian data (Newave deck of May 2017). For that, it reads the "system.DAT" from Newave deck and transformes the data in the format that COPA reads. One assumption was to maintain the transmission line configuration of ONS, which implies working with 5 regions and not more with 4. The fifth region has no load neither generation, but it has bounds on transmission. It will be call “BR11” to follow the ONS nomenclature. The unit of values is Mwmedio (Mwmonth).

The resultant file is named “lineCapacities \_br.csv” and does not consider losses in transmission. It will be considered in GAMS code.

## Creating\_intermittent\_opts\_br

This script reads ANEEL data of wind and solar power plants (existing, being constructed and non-constructed) and prepares a file with max capacity and investment cost for wind and solar power plants.

Assumption: 2 categories of plants for each region:

1) Existing, being constructed and planned: existing, being constructed and non-constructed yet. They will receive investment cost zero and capacity of each region.

2) Non planned: investment cost in each region and a huge potential. The references for these numbers are the Brazilian Atlas of potential. For solar plants, we’ve considered 0.1% of each region area will be used to plants installation.

Concerning the investment costs of wind power plants, we’ve used the EPE document: “Energia Renovável – Hidraulica, Biomassa, Eolica, Solar e Oceanica” from 2015. The monetary actualization was done by IPCA index (to take values to April, 2017) and means currency taxes from IBGE.

Regarding investment costs in solar PV power plants, this script has a plot comparing different PV utility scale costs. It is a good step to compare different references. The chosen one was IRENA: “The Power of Change”, 2016.

## Creating\_hydro\_br

This script creates the hydro files for Brazilian version of COPA. They are br\_shype\_hydro.feather and hydro\_data\_br.csv.

Concerning the br\_shype\_hydro.feather, the script uses the daily inflows for 4 Brazilian regions and converts it to hourly scale, by dividing the daily data by 24 hours.

Regarding the hydro\_data\_br.csv, this file gives the bounds of equivalent reservoirs. The parameters are as follow:

* MaxReservoir: max capacity of storage of each subsystem. This data was achieved in the ONS website. Link is in the script.
* MaxHydPower is the maximum power that could be generated by the subsystem. It came by “CadUsh.csv” from Newave deck.
* MinFlow and MaxFlow are the minimum and maximum flows that can come to the reservoirs respectively. The MinFlow was done by the minimum observed value from the daily inflows from 1979 to 2014. The MaxFlow is simply a very big number.

## Creating\_solar\_feather\_br

This script creates time series of capacity factor of solar PV resource for Brazilian version of COPA. The data is from renewables ninja. I got the information considering 0% of loses in order to obtain directly the capacity factor of the place. While they calculate the production, they consider the capacity factor. If I set losses in the system, the production will be decreased by this percentage of loss. That’s why I put zero in system losses. Concerning the capacity, I set 1 kW. My search summary:

* Period: 2012-01-01 to 2015-12-31
* Capacity: 1kW
* System loss: 0
* Tracking: none
* Tilt: 35 degrees

Azimuth: 180 degrees.

About the data: documentation from the website:

* time: always a UTC date/time stamp;
* output: mean power output over the timestep in kW.

I choose three locations in each region and got information of each year for these locations. I’ve created a new time series that contains hourly values for 2012 to 2015 and the capacity factor of each hour is the mean of the capacity factor of the chosen regions.

## Creating\_wind\_feather\_br

This script creates the time series of capacity factor of wind resource for Brazilian version of COPA. The data comes from Renewables Ninja and the methodology is the same of “Creating\_solar\_feather\_br”.

Search summary:

* Period: 2012-01-01 to 2015-12-31
* Capacity: 1kW
* Hub height: 80 m
* Turbine model: Vestas V80 2000

I’ve chosen three locations in each region and got information of each year for these locations. I’ve created a new time series that contains hourly values for 2012 to 2015 and the capacity factor of each hour is the mean of the capacity factor of the chosen regions.

## Summary

|  |  |
| --- | --- |
| Script | File names |
| Creating\_intermittent\_opts\_br.R | br\_intermittent\_opts.csv and br\_intermittent\_opts.feather |
| Creating\_invest\_opts\_br.R | investOpts\_br\_thermal.sources.csv |
| Creating\_load\_br.R | load\_Br.csv and load\_Br.feather |
| Creating\_lineCapacities\_br.R | linesCapacities\_br.csv |
| Creating\_solar\_feather\_br.R | solar\_GAMS\_br.feather and solar\_GAMS\_br.csv |
| Creating\_wind\_feather\_br.R | wind\_br.csv and wind\_br.feather |
| Creating\_hydro\_br.R | hydro\_data\_br.csv, hidro\_data\_br.feather and br\_shype\_hydro.feather |
| Creating\_load\_br\_adapted.R | Load\_Br\_2014.feather and load\_br\_2014.csv |

## Calibration\_inflows.R

(continue)

## Creating\_shype\_ONS\_2012.R

(continue)

# Analizing results scripts

All scripts of this session need the variable “results” from “load\_data\_write\_gdx.R” to be loaded.

## Analizing COPA Brazil results.R

It make graphs to check key points of COPA Brazil model. The first one contains the renewable generation by source (Wind and Solar PV) by region in MWh. The second one shows the thermal generation by source. The third graph shows investment on PV, thermal and wind by region. It also has an appendix comparing variable costs of thermal power plants.

## Cleaning\_ONS\_validation\_data.R

This script reads ONS website data related to generation on 2012 for each region and cleans it in order to get tidy files to be read by "ONS\_validation\_data\_2012\_complete.R". The ONS data was achieved on its new website: (<http://www.ons.org.br/pt/paginas/resultados-da-operacao/historico-da-operacao>). We run this script four times, changing only the name of the regions and the results are csv tables with generation by technology and for each day of 2012. We´ve created one variable (namesNew) that saves the names of regions and we just have to change this variable in each run in order to get the file of each region.

## Cleaning\_ONS\_transfer\_data.R

This script simply reads ONS file containing electricity exchanges by subsystems, selects relevant information and saves it in a table named “transfer\_ONS.csv”.

## ONS\_validation\_data\_2012\_complete.R

This script compares results of generation of electricity by source and region of ONS and COPA Brazil. Daily values of electricity generation from ONS. Period: 1/1/2012 to 12/31/2012. It reads csv files created in “Cleaning\_ONS\_validation\_data.R” and creates figures comparing technology generation in each region. It also saves these figures in the folder “COPA/runs/Validation\_2012/figures”. This script generates the same results of “ONS\_validation\_data\_region.R”, but in form of a loop. I just have to check with Johannes how can we save the summary plots in one pdf file for each region.

## ONS\_validation\_data.R

This script creates graphs in order to compare electricity generation of ONS and COPA Brazil. It compares the generation of ONS for the period 2012-07-01 to 2012-07-31 for each region.

In “Validation\_2012” folder, we will find four similar scripts: ONS\_validation\_data\_SE, ONS\_validation\_data\_S, ONS\_validation\_data\_NE and ONS\_validation\_data\_N. They create plots comparing hydro, thermal and wind (when it is the case) generation between COPA and ONS for run named “Validation\_2012”, that has the period of 01-01-2012 to 12-31-2012. The main figure of each script is a multiplot with generation by source and points graphs, showing the dispersion of ONS and COPA per source. These figures are saved in the following path “C:\Users\cancella\Google Drive\!IIASA\COPA\runs\Validation\_2012\figures”.

## Creating\_statistical\_indicators.R

This script reads tables from ONS\_validation\_data\_region scripts containing correlation and RMSE between time series of electricity generation for each region and creates a summary indicators table. It also makes plots of each indicator (correlation and RMSE) by region.

## Comparing\_transfer\_ONS\_COPA.R

This script compares ONS daily electricity transfers from 2012 to COPA transfer results. The main result is a .pdf figure containing transfer graph from ONS and COPA side by side. This figure is saved in “COPA/runs/Validation\_2012/figures”.

## Understanding\_hydro\_generation.R

This script compares hydro parameters from “hydro\_data\_br\_2012\_1.csv” with the results of COPA model in order to check the attendance of constraints. It makes a plot of storage levels of each subsystem in the same graph.

Other result of this script is a plot that compares the total hydro and thermal generation between ONS and COPA.

## Comparing\_inflows\_x\_hydro.R

This script compares the inflows from “br\_shype\_hydro.feather” file and the COPA results in terms of hydro generation. It also compares the wind resources of “wind\_br.feather” to wind production in COPA. The main objective is to verify if the model is generating according the availability of these resources.

## Comparing\_COPA\_ONS\_load.R

This script compares the website ONS load with the load that Wilkens sent us. It creates a graph to compare this data. It seems they are very close to each other, but I need to check with Johannes a question regarding the dates. I’m not able to merge the dates of this two time series.

Update comparing\_inflows\_x\_hydro.R

# important decisions

1. **What investment cost values should I consider to the natural gas plants?**

Options: Simple cycle or combined cycle.

Decision: It is a question because we have many technologies of natural gas power plants. There is a 2016’s report made by Planning Energy Company (EPE in Portuguese) that describes the electricity generation technologies and gives some ranges for investment costs of these technologies. Using this document, I obtained 61% of thermal plants coming from combined cycle in Brazil. So I assume that all thermal plants have combined cycle costs.

1. **What reference should I use for the investment costs?**

Options: EPE 2016’s report or Schroder *et al.* (2013).

Decision: I’ve chosen to use EPE’s report, because it is a more recent than the Schroder *et al* (2013) and it considers different ranges for the technologies.

1. **What values of investment costs should I use?**

Options: Under the assumption I’m using the EPE report, my options are the upper bound or the lower bounds for the costs.

Decision: I’ve chosen the upper limits of the technologies costs trying to be more conservative, but I don’t have much sensitivity about the costs values. Maybe a good idea would be making scenarios afterwards changing some key variables in a range (the costs is one of them) to analyze model’s behavior.

1. **What currency should I use?**

Options: U$ from May 2017 or R$ from May 2017.

Decision: My variable costs are in R$/MWh, so I though the best currency would be R$ from May 2017. The monetary restatement was made by applying the average exchange rate for the year in question to the required values, ie the average exchange rates for 2013, 2015 and 2016 were applied. Afterwards, we have the values ​​in Brazilian reais (R$) of 2013, 2015 and 2016. Subsequently, the IPCA (inflation official index in Brazil) correction was applied to reach the values ​​of May 2017.

1. **Available capacity of non-existent thermal plants.**

Options: I can use the nominal capacity like a proxy of the available capacity, or I can try to estimate the available capacity in some way.

Decision: I don’t have information about interruption of these plants in Newave deck. Because of that, by now, I simply used the maximum capacity as the available capacity for these plants. I’m also considering to estimate it using data from the existent plants. However it is not implemented yet.

1. **Future issue: Thermal plants fuel addition**

Options: The Swedish version of the thermal power plants file doesn’t specify the fuel of each plant. I would like to describe it for the Brazilian version.

Decision: I’ve decided to prepare the file with fuel information, but I`ve not prepared the R and GAMS codes to receive this modification yet. I guess Johannes can help me with it. In the meanwhile, I’m working on the preparation of others input files.

1. **R issue: Why some columns of “double” are transformed into “factor” when we read.csv files?**

Decision: Sometimes when I read a .csv file that I’ve saved in a previous script, some columns of my tables come like “factor’ and it is a problem because I’m not able to make calculations with this kind of variable.

The way I found to solve it was to apply the following line:

our\_table[sapply(our\_table,is.character)]🡨apply(our\_table[sapply(our\_table,is.character)],function(x)as.numeric(as.character(x))).

Is there a better way to deal with it?

1. **Modification on script functions\_gdx\_transfer.R**

Around line 47 of this script, I realize the load file was created based in feather file. Because of that, my load was not being saved. I changed the file for the .csv and changed the reading function. It was “read\_feather” and now it is “read\_delim”. I saved the initial file as “functions\_gdx\_transfer - original.R” and the new one as “functions\_gdx\_transfer\_br.R”. Important: by the time we will make the runs, we need to change the name of this script to R functions read it properly. The name must be “functions\_gdx\_transfer.R”.

# To do

1. Adjusting the thermal power plants costs file: do not multiply for capacity – OK
2. Pass to R monetary actualization - OK
3. Adjust transmission capacities file: ONS configuration and to consider last month of 2017 like bounds. – OK
4. Intermittent file costs. – OK
   1. Search for more recent values of investment costs.
5. Hydro\_data\_br
   1. Update the MinFlow column. – OK
   2. MaxHydPower: aggregation of hydro power plants from CadUsh.csv by subsistem. Be aware about the number of turbines in each plant. - OK
6. Renewables capacity factor files: renewables ninja. - OK
7. Run tests: period one month.

Check the scripts of capacity factor files for wind and solar. – OK

Documentation of scripts starting by creating\_hydro. – OK

More recent investment costs of intermittent renewables (manly PV) – OK

Review of script creating\_inttermitent\_opts\_br - OK

Check all the input files and scripts OK

To separate data treatment from the construction of the file on creating\_intermittent\_opts\_br. OK

Load file period 2012 to 2014 OK

Run tests: period one month. OK

More recent costs PV and wind. OK