Description of the data used in the PyPSA Vietnam paper

Minh Ha-Duong[a]

2019-08-14

1 Introduction

An Optimal Power Flow (OPF) model solves the problem of determining the best operating levels for electric power plants in order to meet given demands throughout a transmission network, usually with the objective of minimizing operating cost. It is a non-linear optimization problem, combining the economic dispatch and the power flow problem. A typical load dispatch center solves the OPF hourly.

Alexander Kies from University of Frankfurt kindly communicated us the dataset of the OPF model named PyPSA used for their renewable energy integration study in Vietnam [1]. It was given to us without the raw data or the scripts. However, since PyPSA is open source and actively developed, we are examining the opportunity to use it in VIET.

To this end, we need to understand the kind of data it needs. This note analyzes the dataset we were given. It relies on reading the online documentation and reading the paper.

The dataset comprises 11 files. The format is straightforward, PyPSA relies of Python pandas dataframe library. The files are comma separated value format, ASCII text, some with very long lines.

There are two kind of tables.

- Those describing components of a given type (lines, generators...). Each component has a row, and the columns describe its attributes. This works for scalar attributes, which can be either: empty; a Float using the dot as decimal separator; an unquoted Strings; or the boolean string True. In this dataset nothing is False, no integer, no NaN some floats are scientific notation, some floats could be rounded in decimal.
- For attributes that are time-dependent (solar irradiation, water inflow….), the table list the components in columns, and the time dimension in rows (one line per snapshot)

We look at the 11 files in four categories: model structure, grid, supply, demand.

2 Model structure

2.1 network.csv

This file defines the PyPSA model version 0.13.2 , the geographical system srid 4326, and leaves the name empty.

2.2 snapshots.csv

- name — a time point at which to compute the system state
- weightings — the duration (in hours) represented by snapshots

a Directeur de recherche CIRED/CNRS & Vietnam Initiative for the Energy Transition. minh.haduong@vietse.vn

The model has hourly timepoints for two years, that is 17544 = 8772 * 2, from 2015-01-01 00:00:00 to 2016-12-31 23:00:00.

2.3 carriers.csv

- name
- marginal
- capital
- co2_emissions (tonnes / MWh)
- short_name

The table defines twelve carriers: bioenergy, hard coal, hydro, ocgt, nuclear, oil, wind , solar, perpetuum, battery, hydrogen, hydro ror

However, carriers battery and hydrogen are not used/mentioned anywhere else in the dataset.

Carrier perpetuum is a backstop technology at high capital cost.

Attributes marginal and capital are not specifically defined in the online documentation, it only says that attributes relevant for global constraints can be stored in this table.

3 Grid components

3.1 buses.csv

- Name
- Area (left blank in current dataset)
- Area_offshore (left blank in current dataset)
- Country (VN)
- Nominal voltage (kV)
- Latitude, longitude (using coordinate system WGS84 / SRID 4326)
- substation_lv (always set to True in the current dataset)

The current dataset has 46 bus, all 500kV, AC.

The meaning of the attribute substation_lv is not known. Adding a Province attribute might be useful, even if it can be obtained from coordinates.

3.2 lines.csv

- Bus0
- Bus1
- Length (km)
- Type (The kind of line, a string from a list of industry standards)
- num_parallel (number of parallel lines)
- [Shunt susceptance b (Siemens)]
- [Series resistance r (Ohm)]
- [Series reactance x (Ohm)]
- s_nom Limit of apparent power which can pass through branch (MVA)
- s_nom_extendable True if Optimal Power Flow solver can extend the line limit
- Capital_cost Capital cost of extending s_nom by 1 MVA.

The current dataset as 65 lines, all declared of type 490-AL1/64-ST1A 380.0. The num_parallel is left to the default value of 1.

Since Type is defined, line electrical parameters for r, x and b are redundant: PyPSA recomputes them from the length of the line and industry standards. The current dataset do not define the limit of apparent power s_nom but let it to default value (zero) and let the solver to computes an optimal limit given a capital cost.

4 Supply components

4.1 generators.csv

- Name
- Bus
- Carrier
- Component ("PP" or empty)
- p_nom (Nominal power in MW, used for limits in OPF)
- max_hours (0.0 or empty)
- p_nom_extendable (True to allow capacity p_nom to be extended in OPF).
- p_nom_min If p_nom is extendable in OPF, set its minimum value in MW
- p_nom_max If p_nom is extendable in OPF, set its maximum value in MW
- marginal_cost Marginal cost of production of 1 MWh.
- capital_cost Capital cost of extending p_nom by 1 MW.
- weight

The table has 414 lines = 46 bus x 9 carrier. It has two parts.

(1) Existing power plants connected to a bus, represented by row with component == "PP", p_nom and p_nom_min are non-zero and equal.

(2) Potential generators: represented for all other couple (bus, carrier). Column component is left empty, p_nom and p_nom_min are zero, p_nom_max is infinite for fossil fuel technologies or defined according to a natural resource potential model (not given to us) for wind, solar and hydro_ror.

The meaning of max_hours is not described in the doc.

Carriers hydro, battery and Hydrogen are not used in this table. See storage_units.csv for hydro.

The dataset sets p_nom_extendable to always True, and let the OPF determine installed capacity, subject to p_nom_max constraint.

4.2 generators-p_max_pu.csv

The time dependent attribute p_max_pu represents the maximum power generation coefficient (instantaneous capacity factor), for the intermittent generators {solar, hydro ror, wind}. The table has 138 columns: 46 nodes x 3 carriers and one row for each time point (see snapshot.csv)

The contents is the maximum output for each snapshot per unit of p_nom for the OPF. The p_max_pu for other kind of generators are left to default: 100% of p_max can be dispatched all the time.

Probably defined according to some natural resource potential model (not given to us).

4.3 storage_units.csv

- name
- bus
- carrier (hydro for all components)
- component Either “Store” or empty
- p_nom Nominal power (MW) for limits in OPF, or 0.0 if none existing
- max_hours Maximum state of charge capacity in hours at full output capacity p_nom
- p_nom_extendable (True for all rows, allows OPF solver to invest in capacity expansion)
- p_nom_min Existing capacity (MW)
- p_nom_max Technical potential (MW)
- marginal_cost Marginal cost of production of 1 MWh.
- capital_cost Capital cost of extending p_nom by 1 MW.
- efficiency_store Efficiency of storage on the way into the storage.
- efficiency_dispatch Efficiency of storage on the way out of the storage.

The table has 46 lines, one per node. 22 have non-zero p_nom.

4.4 storage_units-inflow.csv

The time-dependent attribute inflow represents Inflow to the state of charge (in MW), e.g. due to river inflow in hydro reservoir.

Table has one row per time point, one column per storage unit (=46, one storage unit per bus).

Defined with MERRA model according to the manscript.

5 Demand components

5.1 loads.csv

There is one load component per bus, named identical.

5.2 loads-p_set.csv

The time-dependent attribute p_set represents the active power consumption (MW).

Table has one row per time point, one column per load (=46, one load component per bus).

6 References

1. Dinh Thanh Viet *et al.* A cost-optimal pathway to integrate renewable energy into the future Vietnamese power system. in *GSTD 2018 proceedings* 144–149 (IEEE, 2018).