# Minimum data requirements for SWITCH modeling

The following list contains the minimum data requirements for developing and running a SWITCH simulation for a power system. It assumes previous basic knowledge of the model concepts, which can be reviewed in Nelson et. al (2012) and Fripp (2008).

## Load Zones

SWITCH’s basic unit of analysis is load zones, which represent portions of the grid that are not transmission constrained in present time. Even though load zones are defined by electrical terms, they need to have a geographical representation. Being the basic unit of analysis, all elements being simulated have to be linked to a certain load zone in order to be included in the simulation.

Therefore, the following data is needed:

* GIS data from the territory where the simulated system is located. This is usually in the form of shapefiles depicting administrative boundaries, such as states, counties, cities, etc. However, load zone polygons can be customized not following any particular existing division.
* Supporting information for segmenting the grid into non-constrained areas. In the case of Chile, for example, every month the load dispatch center identifies subsystems within the grid whose marginal costs are decoupled due to transmission constraints between them. Then, each relevant bus is assigned to a specific subsystem, thus creating load zones.

## Grid data

SWITCH requires knowing the current state of grid data, including:

* All relevant buses and lines above the distribution level, this is, usually over 23 kV or on larger systems over 110 kV.
* For lines, include length, capacity, resistance, reactance, number of circuits (for N-1 operations) and if it’s operated under N-1 constraints. Include also the bus or substation of origin and end.
* GIS information on the main buses location is needed. It is useful to have GIS information of the lines, maybe on a shapefile containing the grid layout.
* Lines that connect load zones should be identified, indicating which load zones they connect.
* Spinning reserve constraints should be provided. They can be at the load zone level and/or depending on certain generation technology.

## Generation

For existing generation, the following information is needed:

* For all plants, geographical location is required, as well as the bus they connect to in the grid. If no GIS information is available, a suitable way to identify to which load zone they belong must be derived.
* Carbon intensity, in ton(CO2)/MWh, should be included also for each plant that emits it. Alternatively, carbon intensity can be provided for each type of fuel used in ton(CO2)/MMBTu
* For thermal plants, identify the fuel and technology they use, maximum and minimum capacity, heat rate in MBTU/MWh, and whether it is cogenerating.
* For hydropower plants, separating between run of river, pumped hydro and reservoir is required. Maximum and minimum capacity is also needed as well as capacity factors with the highest resolution possible (hourly ideally). These should be ideally reported for dry, normal and wet years, as it critically affects their output. We usually require data for at least a year, but 3-5 years is preferred if available.
* For wind and solar plants historical hourly capacity factors are needed, as well as maximum and minimum capacity. In lieu of this information, hourly meteorological information (i.e. wind speed and radiation) for the same sites can be used to estimate capacity factors using appropriate modeling tools. We usually require data for at least a year, but 3-5 years is preferred if available.
* All other relevant plants need to be included, mentioning their technology (biomass, waste, biogas, etc), capacity and location. To assess capacity factors at least a yearly average is required, but higher resolution capacity factors would be required for plants whose supply has an inherent variability.

For future generation projects, the following information is needed:

* If there is a national or regional level master plan that includes future plants that will be built, then they can be included following the guidelines mentioned before for existing plants. They should include the start (installation) year as well.
* To let SWITCH assess the suitability of certain developments, specific data can be provided. All sittings should be identified by their spatial coordinates, clearly indicating projection used.
	+ Siting for solar and wind projects. This should include at least 1 year of hourly radiation and speed measurements, to assess capacity factors for these projects.
	+ Siting for hydropower projects. This should include river flow measurements, at least at the monthly level, in order to assess seasonal variation of capacity factors. Rainfall information for a number of years is also needed, so river flow can be parameterized by those in order to simulate specific dry, normal and wet years. However, it’s useful to inform specific reservoir/RoR siting with head and expected flow/capacity when available, including the year they would be connected.
	+ Forbidden siting for thermal plants due to pollution saturation. These usually take the form of geographical polygons where locating thermal plants, especially coal, is not allowed.
	+ Siting for biomass projects. This can take many shapes, such as identifying specific biomass source points or relevant forests. Supply curves are relevant to assess incremental costs for generating with this source, as well as seasonal variability that affects temporal behavior of capacity factors.
	+ Siting for geothermal projects should include expected capacity, as it’s quite difficult to parameterize the size of a plant based on geological data such as temperature, steam/water flow, among others. Sittings that allow a standard 40-50 MW/90% capacity factor module plant should be provided (these figures can vary, of course).
	+ Siting for marine energy projects should be treated similarly to geothermal sites. Studies that support generating with this source should be provided.
	+ All siting should include some sort of GIS data, or at least be assigned to a specific load zone in advance.

## Demand

Demand data is required at both the historical and the projected level, as follows:

* Historical hourly demand, for at least one year, should be obtained at the load zone level. Usually, demand would be informed at the bus level, so an assignment will need to be constructed to carry this to the load zone unit. This historical demand is used to build an annual hourly profile that is later applied to the future energy demands, so an hourly projection can be built.
* It is of utmost importance that all hourly data is correlated. This is critical to correctly assess the viability of solar and wind energy meeting demand, especially in peak hours.
* Projected annual energy demand for each load zone is required, up to the end year of the study (usually between 20 and 50 years into the future). Ideally, these projections should come from an official source, such as the local energy agency/ministry or dispatch center. The model does use hourly future demand. We usually use past hourly profiles and apply them to future annual energy forecasts. However, if there are hourly demand forecast models available we would use those results to feed into SWITCH.

## Costs and prices

There are a number of different cost and price data that needs to be informed:

* Annualized sunk cost for distribution and transmission, including investment and O&M expenses, at the load zone level. Units are US$.
* Expansion costs for distribution and transmission, at the load zone level.
	+ For distribution estimate the cost of expanding 1 MW of capacity at that level in US$.
	+ For transmission estimate the cost of expanding 1 MW-km of grid at the most usual voltage levels in US$. That figure must include substations also, so a method must be derived to add those costs up.
* Current and future annual fuel prices, up to the end year of study, must be provided. They can be different per load zone, if for any reason there are specific market conditions that justify so. Units should be in US$/MBTU.
* Generation costs, which include non-fuel variable costs, fixed O&M costs and overnight construction costs. These can be for each existing generation unit or for a particular technology.