## Plan4res input data

This chapter presents the conventions deployed to name the files, the format of datasets, and the main transformations.

The plan4res dataset is composed of 1 excel file and 1 optional csv file, plus a serie of CSV files for the scenarised timeseries. The excel file contains the following sheets :

* Parameter : main parameters of the run
* ZP\_ZonePartition : describes the geography
* IN\_Interconnections : data related to interconnections
* ZV\_ZoneValues : data related to coupling constraints, in particular electricity demand
* TU\_ThermalUnits : data related to units modelled as ‘thermal’ in SMS++
* SS\_SeasonalStorage : data related to seasonal storage
* STS\_ShortTermStorage : data related to short term storage and other flexibilities modelled as short term storage in SMS++
* RES\_RenewableUnits : data related to units with uncertain profile (mainly PV, Wind and RunOfRiver)
* TS\_HourlyTimeSeries : contains the deterministic time series. This file is optionnal as it can be replaced by a csv file containing the same data. This is usefull when the number of deterministic timeseries is large (eg more than 50)

The CSV files containing scenarised time series follow the common format on time series, ie 2 lines of heading (one with global naming information, the other with the names of the time series); The first column contains the UCT timestamp (yyyy/mm/dd hh:mm) and the following columns are the different scenarios of the current timeserie=> one CSV file contains one scenarized timeserie. Scenarios are stamped by the value in the line2 (usually years, eg 1970, 1971, 1972…);

### Description of geography

Each dataset is linked to a geographical area that may be partitioned. Different partitions may be used for dealing with different levels of constraints or computations.

Those partitions are described in files or sheets named ZP\_ZonePartition, that follow the same rules:

The first column ‘Level1’ lists the lowest level partition. The partition ‘level2’ is a higher level partition, meaning that the regions in Level2 are larger than the regions in level1, and each region in level2 is composed of a list of regions in level1. Level3 partitions again are bigger than level2. At each level we may have different partitions obtained by regrouping regions differently (here we have 2 different ways of grouping the regions of level2: level3part1 and level3part2).

* Level1 = identifier of the first level zone (e.g., datazone) - string
* Level2= identifier of the second level zone (e.g., cluster) - string
* Level3Part1 = identifier of the third level zone (e.g., region), in a first partition – string
* Level3Part2 = identifier of the third level zone (e.g., region), in a second partition – string
* Etc.

A row with values L1, L2, L31, L32 … means that zone L1 belongs to second level zone L2, and this zone belongs to third level zones L31 and L32, …; Each L1 belongs to a unique L2; Each L2 belongs to a unique L3 for the first partition, and to a unique L3 – that can be different- for the second partition, etc…

The Figure 1 below illustrates this.

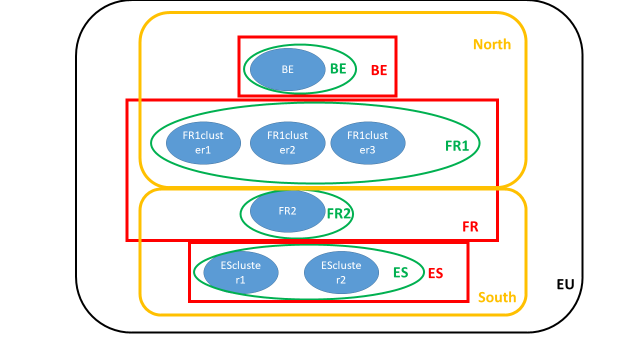


Figure 1: partitions

In Blue: Level1; in Green: level2, in red and orange: level3 (partition1 and partition2), and in black level4. Here there are 7 zones in level1, that are aggregated in 4 level 2 clusters: BE, FR1, FR2 and ES. Each level1 zone belongs to a unique level2 cluster. At level3 there are 2 partitions: Level3part1 partitions the 4 clusters as North=BE+FR1, South=FR2+ES while Level3part2 partitions the same 4 clusters as BE=BE, FR=FR1+FR2 and ES=ES.

### Excel sheet Parameter

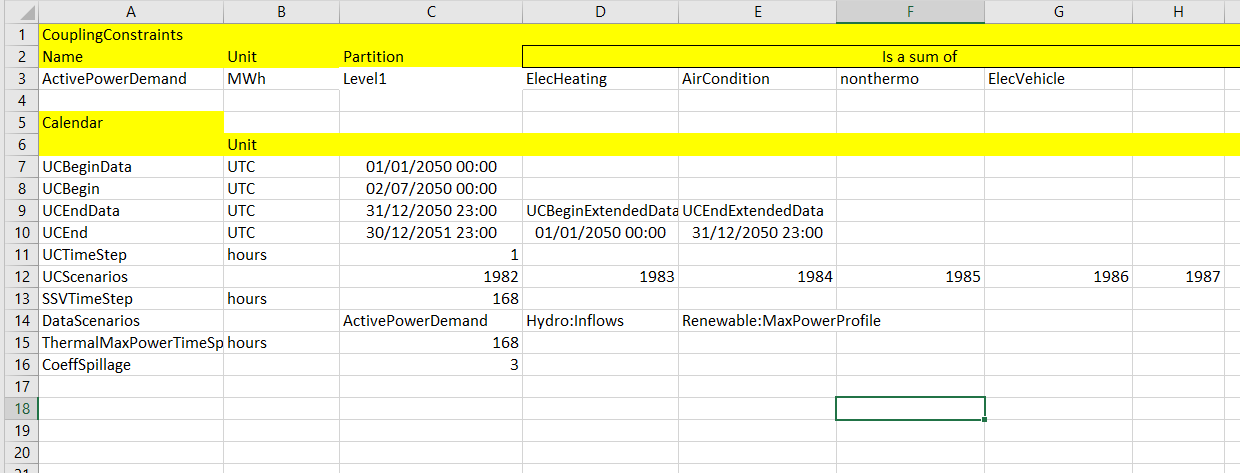


Figure 2: Sheet 'Parameter'

This sheet is composed of the following sets of data:

#### Coupling Constraints

this lists the different coupling constraints that may be used. If a coupling constraint is not used, the line is deleted. For each constraint, the Unit is given as well as the Partition, which describes the zones attached to each coupling constraint. The following constraints may be included :

* ActivePowerDemand: this is related to the equilibrium constraint Active power = Active Power demand at each node of the network. The Active Power Demand is computed as the sum of different time-series that are listed in the corresponding line (this means that the power demand is computed as the sum of the electric heating demand, the electric cooling demands….)
* PrimaryDemand: this is related to the primary reserve, in each “zone” (can be a country….)
* SecondaryDemand: this is related to the secondary reserve.
* InertiaDemand: this is related to the inertia requirement.
* PollutantBudget: this constraint is not used in the current version of the tool. It is related to the maximum amount of polluting emissions, for a specific pollutant.

#### Time Parameters

Time is discretized in:

* Time Sets (usually 1 week)
* Time Steps (usually 1 hour)

The granularity of the data in the different groups of TimeSeries may be different. In practice, there may exist hourly, daily or weekly time series. If the time step of the given data is different than the time step of the variable to be filled, it means that:

* If the unit is % or MW/GW, the value of the variable at lower timesteps will be the one that is in the dataset.
* If the unit is MWh/GWh, the value of the variable at lower timesteps has to be adapted (eg 24GWh on a daily timestep means 1GWh on an hourly timestep)

The time horizon is described in by the following variables in the ‘calendar’ block of the Sheet Parameter. These parameters give both the periods where time dependent data are available and the period when we want to run the model. Usually the data (for the timeseries) used are available only for eg 1 year and the user wants to run the model on eg. 1 ½ year. This is made possible by extending the time series, which is made by copying the beginning of the timeserie at the end of the timeserie.

* UCBegin: is the first timestep of the study. It has to be provided as a UTC timestamp.
* UCEnd is the last timestep of the study
* UCBeginData : first timestep of the data used
* UCEndData : last timestep of the data used
* UCBeginExtendedData : first timestep of the available data to be used to extend the dataset for the run
* UCEndExtendedData : last timestep of the available data to be used to extend the dataset for the run
* UCTimestep is the timestep duration. If ‘hours’ is specified in the unit column, it means that it is a number of hours. This means that the total number of time steps of the whole Time Horizon is (UCEnd-UCBegin)/UCTimeStep (in the example below, we have 3 days and a timestep of 1 hour then 72 timesteps. Each time a profile is used, only the data comprised between UCBegin and UCEnd are used
* SSVTimeStep gives the duration of the Time Sets (which is the seasonal storage timestep). In the current example, it is 24 hours (1 days) which means that there are 3 Time Sets of 1 day each. SSVTimeStep has to be at least bigger than UCTimeStep.

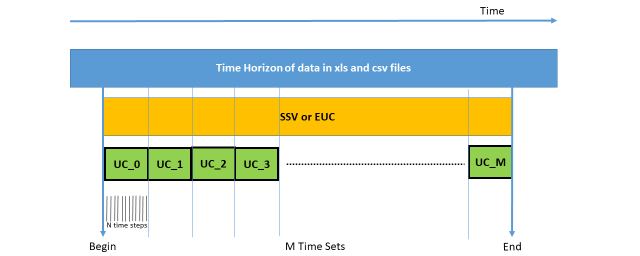


Figure 3: Time Management

#### Scenarios parameters

The UC solver (and the UC simulator) are deterministic, which means it runs on 1 scenario only. The line UCScenario contains the list of scenarios identifiers that will be used for generating the scenario data (not yet included). Here in the example we have 3 identified scenarios: 1982, 1983, 1984. In the data we may have some scenarized timeseries (in our example only the timeseries from the csv file), and some non scenarized timeseries. Scenarized timeseries are identified by a scenario tag (here it is the historic year that was used for computing the data).

When using a deterministic mode, only the first scenario will be used.

The parameters related to scenarios in Parameter sheet are the following:

* UCScenarios is the list of scenarios identifiers. Those identifiers have to be at the top of the column related to this scenario, for all variables of the current scenario.
* DataScenario is the list of variables that are scenarised in the current run. This list can comprise:
  + ActivePowerDemand
  + Thermal:MaxPower (the maximum power for thermal plants, accounting for unavailabilities schedules)
  + Hydro:Inflows (the hydraulic inflows at seasonal reservoirs)
  + Renewable:PmaxProfile (the maximum capacity of renewables)

#### Model parameters

These are parameters used for creating the dataset :

* ThermalMaxPowerTimeSpan (hours) : duration for which the availability of thermal plants is fixed (in the example, the availability is given per week)
* CoeffSpillage : size of the fictive spillage reservoir which is created associated to each reservoir to allow spillage. The size of the spillage reservoir is the size of the main reservoir multiplied by this coefficient.

### Excel sheet ZP\_ZonePartition



Figure 4: Sheet 'ZP\_ZonePartition'

The sheet ZP\_ZonePartition describes the different partitions that are used for dealing with different coupling constraints.

The mapping between the coupling constraints and the partitions is defined in the first block of the sheet parameter: In our example, ActivePowerDemand is using level1. This means that level1 is the level to be used to generate the nodes of the model. In the example we have 14 nodes. In practice the model will receive the constraint at each node ActivePower in the current node=ActivePowerDemand of the current node ; InertiaDemand uses usually level2 (which in the example is Europe). This means that there is only one InertiaDemand constraint ; In practice the model will receive the constraint Total inertia (ie sum of the inertia in all nodes)=inertiademand

### Excel sheet ZV\_ZoneValues

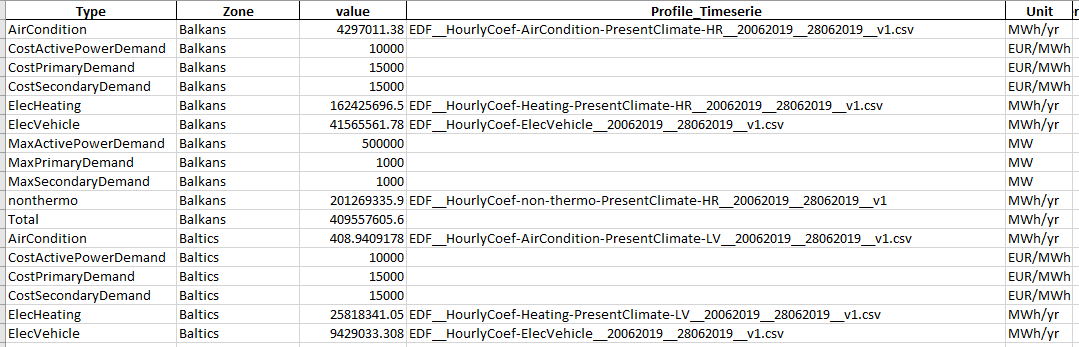


Figure 5: ZV\_ZoneValue sheet

This sheet contains the values of all coupling constraints, as well as the costs associated (imbalance costs), and the names of the scenarised timeseries. In the example the AirCondition timeserie used to compute the ActivePowerDemand for ‘Balkans’ is computed as the value ‘4297011.38’ multiplied by the (usually hourly) values of the timeserie ‘EDF\_\_HourlyCoef-AirCondition-PresentClimate-HR\_\_20062019\_\_28062019\_\_v1’

### Excel sheet IN\_Interconnections



Figure 6: IN\_Interconnections sheet

This sheet describes the characteristics of the lines that are linking the nodes of the network, where the nodes are those described in the partition linked to the ActivePowerDemand variable:

* Name (optional): name of the line (used for processing results)
* Type: type of the line (not used)
* Direction : unused, all lines are bidirectional. To include not bidirectionnal lines, Min or MaxPower =0
* StartLine and EndLine must be nodes defined in the partition linked to the ActivePowerDemand
* MaxPowerFlow and MinPowerFlow are the bounds on the flows for this line
* Impedance (used in the version with AC)
* MaxAddedCapacity and MeanAddedCapacity are the bounds for the Capacity Expansion Model
* InvestmentCost is the cost for extending the line capacity by 1 unit

### Excel Sheet TU\_ThermalUnits



Figure 7: TU\_ThermalUnits sheet

This sheets gives the characteristics of all thermal power plants.

It contains the following data:

* Name
* Zone
* NumberUnits : number of units of the same type at the same location
* MaxPower
* MinPower (optional, 0 by default)
* Pauxiliary: Power taken from the system when off. Optional (0 by default)
* FixedCost: Optional (0 by default) ;
* VariableCost: proportional cost; Optional (0 by default).
* Quadterm: quadratic cost; Optional (0 by default)
* StartUpCost: Optional (0 by default)
* MinUpTime: minimum duration when the plant is on; Optional (1 by default)
* MinDownTime: minimum duration when the plant is off; Optional (1 by default)
* Inertia: max inertia that can be provided by a unit; Optional (0 by default)
* PrimaryRho: this parameter, multiplied by Maxpower, gives the maximum primary reserve that can be provided by a unit; Optional (0 by default)
* SecondaryRho: this parameter, multiplied by Maxpower, gives the maximum secondary reserve that can be provided by a unit; Optional (0 by default)
* DeltaRampDown: maximum gradient when the power is decreased from one time step to the other. Optional (MaxPower by default)
* DeltaRampUp: maximum gradient when the power is increased from one time step to the other. Optional (MaxPower by default)
* UnavailabilityRate: rate of unavailability for failure. Optional (0 by default)
* MeanUnavailabilityDuration: mean duration of unavailability. Optional (0 by default)
* MaintenanceRate: rate of unavailability for maintenance. Optional (0 by default)
* CO2 : emissions per MWh (optional, default 0)
* MaxAddedCapacity and MeanAddedCapacity are the bounds for the Capacity Expansion Model
* InvestmentCost is the cost for extending the line capacity by 1 unit

### Excel sheet SS\_SeasonalStorage

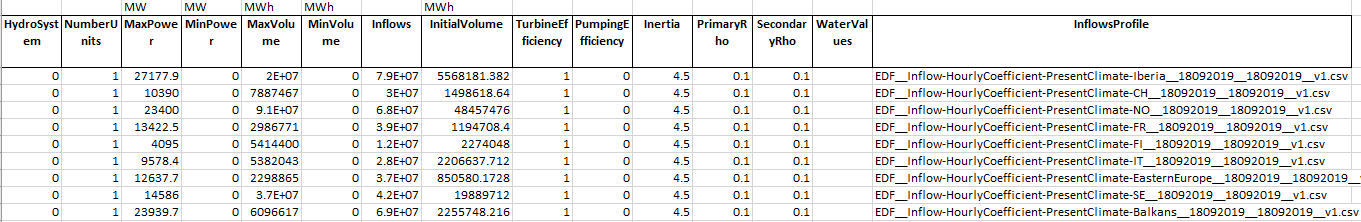


Figure 8: SS\_SeasonalStorage sheet

This sheets gives the characteristics of all seasonal storages.

It contains the following:

* Name
* Zone
* NumberUnits : number of units of the same type at the same location
* MaxPower (MW)
* MinPower (MW, optional, 0 by default)
* VariableCost (€/MWh: proportional cost. (optional, 0 by default)
* DeltaRampDown: maximum gradient when the power is decreased from one time step to the other. Optional (MaxPower by default)
* DeltaRampUp: maximum gradient when the power is increased from one time step to the other. Optional (MaxPower by default)
* MaxVolume (MWh)
* MinVolume (MWh, optional, 0 by default)
* TurbineEfficiency: (optional, 1 by default). This value, multiplied by the flow, gives the generated power.
* PumpingEfficiency: (optional, 1 by default). This value, multiplied by the flow, gives the generated power.
* Inflows: (MWh ; optional, 0 by default). Inflows to the upstream reservoir (energy per year).
* Inflows profile: (optional): time serie profile for Inflows. Multiplied by the Inflows in energy, gives the inflows time serie.
* InitialVolume: (MWh, optional, 0 by default). Initial Volume of the upstream reservoir
* Inertia: max inertia that can be provided by a unit; Optional (0 by default)
* PrimaryRho: (%) this parameter, multiplied by Maxpower, gives the maximum primary reserve that can be provided by a unit; Optional (0 by default)
* SecondaryRho: (%) this parameter, multiplied by Maxpower, gives the maximum secondary reserve that can be provided by a unit; Optional (0 by default)
* WaterValues: Optional; Used if external water values are used in the current run. Contains the name of the file/sheet where the water values are stored. When using the simulation mode with 1 BV file for all units, this file is only given in the first line ; when using the simulation mode with 1 BV per unit, this file is given at each line ; in optimisation mode, this is not required (output of the model)

### Excel sheet STS\_ShortTermStorage

This sheet is used for :

* Pumped storage
* Batteries
* Other flexibilities such as residential flexibilities or EV, which are modelled as short term storages



Figure 9: STS\_ShortTermStorage sheet

It contains the following data:

* Name
* Zone
* NumberUnits : number of units of the same type at the same location
* WindowsSize (hours, optional): size of the windows where load shifting is allowed
* MaxPower (MW) ; can be a (deterministic) timeserie, in this case : name of the time serie
* MinPower (MW, optional, 0 by default) ; can be a (deterministic) timeserie, in this case : name of the time serie
* Cost: proportional cost. (€/MWh, optional, 0 by default)
* MaxVolume (MWh) ; can be a (deterministic) timeserie, in this case : name of the time serie
* MinVolume (MWh, optional, 0 by default) ; can be a (deterministic) timeserie, in this case : name of the time serie
* TurbineEfficiency: (optional, 1 by default). This value, multiplied by the flow, gives the generated power.
* PumpingEfficiency: (optional, 1 by default). This value, multiplied by the flow, gives the generated power.
* Inflows: (optional, 0 by default). Inflows to the upstream reservoir.
* InitialVolume: (MWh, optional, 0 by default)
* VolumeLevelTarget: (MWh, optional) used to force the optimisation to reach this volume at the end of each time set. If there is a VolumeLevelTarget filled, then the minimum volume constraint is replaced by this value at first and last timesteps of each time set.
* MaxPrimaryPower: (MW) maximum primary reserve that can be provided by a unit; Optional (0 by default)
* MaxSecondaryPower: (MW) maximum secondary reserve that can be provided by a unit; Optional (0 by default)
* MaxAddedCapacity and MeanAddedCapacity are the bounds for the Capacity Expansion Model
* InvestmentCost is the cost for extending the line capacity by 1 unit

### Excel Sheet RES\_RenewableUnits



Figure 10: RES\_renewableUnits sheet

This sheets gives the characteristics of all renewable units: windpower, PV power and run-of-river.

It contains the following data:

* Name
* Zone
* NumberUnits: number of units of the same type at the same location
* MaxPower (MW for PV and WindPower ; for RunOfRiver MWh/year)
* MinPower (MW, optional, 0 by default)
* Inertia: max inertia that can be provided by a unit; Optional (0 by default)
* Gamma (optional, 1 by default): this parameter is used by the model to determine the maximum available primary and secondary reserve. It is used to take into account the fact that some renewable units, due to the uncertainty in their maximum capacity, may not be able to provide reserve at full capacity.
* MaxAddedCapacity and MeanAddedCapacity are the bounds for the Capacity Expansion Model
* InvestmentCost is the cost for extending the line capacity by 1 unit

### Excel Sheet SYN\_SynchCond



Figure 11: SYN\_SynchCond sheet

This sheets gives the characteristics of all synchronous condensers. Synchronous condensers are units that consume energy from the grid and provide inertia.

It contains the following data:

* Name
* Zone
* NumberUnits: number of units of the same type at the same location
* MaxRotatingConsumption: this is the maximum power of the unit, that being consumed by the unit, will allow it to provide inertia.
* ConstTerm: Optional (0 by default) ;fixed cost
* Inertia: max inertia that can be provided by a unit; Optional (0 by default)

#### Case of Investments

For running the capacity expansion model, 3 clumns need to be added to the sheets corresponding to assets for which investment is possible :

* MaxAddedCapacity (MW) : this is the maximum capacity that may be added
* MaxRetCapacity (MW) : this is the maximum capacity that may be taken out
* InvestmentCost (€/MW) : this is the cost for investing into 1MW of the given capacity in the given region
* DecommissionCost (€/MW) : this is the cost for taking out 1MW of the given capacity in the given region

These columsn may be added to the sheets :

* IN\_Interconnections
* TU\_ThermalUnits
* STS\_ShortTermStorage
* RES\_RenewableUnits

### Excel Sheets TS\_xxx (time series)

TS sheets contain deterministic time series with different time granularity. 3 different sheets can be used (see below). The hourly time series sheet can be replaced by a csv file (see format below

* TS\_HourlyTimeSeries, for hourly data:

Figure 12: TS\_HourlyTimeSeries sheet

* TS\_DailyTimeSeries, for daily data. In that case the convention in the model is that in case of data in MW (or GW…), the data will be duplicated at each timestep of the day (if timesteps are lower than 1 day); In case of MWh (or GWh…), the data has to be either associated to the first time step of the day (and 0 at the others), or it has to be converted to the duration of the timestep and then duplicated.



Figure 13: TS\_DailyTimeSeries sheet

* TS\_WeeklyTimeSeries: for weekly data. The same rules apply for MW/MWh data than for daily data.



Figure 14: TS\_WeeklyTimeSeries sheet

### Deterministic timeseries csv file

This file may replace the TS\_HourlyTimeSeries sheet. It contains a column per timeserie and a row per hour.



Figure 15: deterministic time series CSV file

Most variables may be given as deterministic timeseries.

### Scenarised Time Series

For all scenarised data, the corresponding time series are identified in the excel file (eg in ZV\_ZoneValue, in the Profile\_Timeserie column. Time series identifiers must be either a single name, or a csv file (ie nametimeserie / nametimeserie.csv). In the case of a single name, the timeserie will be found either in the TS\_xx sheets or in the deterministic time series csv file. In case of ‘name.csv’, each csv file, corresponding to each timeserie, will be found in a specific directory (the path of this directory is an input to the plan4res formatting tool).

Scenarised timeseries can be :

* Electricity demand time series, or components of the electricity demand time series (for the different part of the electricity demand such as EV, heating, cooling, ….)
* WindPower and PV profiles
* Run-of-river profiles
* Inflows to seasonal reservoir profiles
* Maximum power of ‘thermal’ generation

#### Time Series for Inflows

Inflows at seasonal reservoirs (weekly profile time series at country geographic granularity) are given in the following files:

* + EDF\_\_Inflow-HourlyProfile-ScenarioName-CT\_ZoneCode\_\_ddmmyyyy\_\_ddmmyyyy\_\_vi.csv is the profile time serie of the inflows to seasonal storages (in MW)

Inflows time series in MW are provided with a weekly timestep. The value has to be duplicated on each hour of the week.

#### Time Series for Water Values

This optional time series contains water values that will be used in the simulation, if not already computed by plan4res. It has to be a .csv file, with the following format :

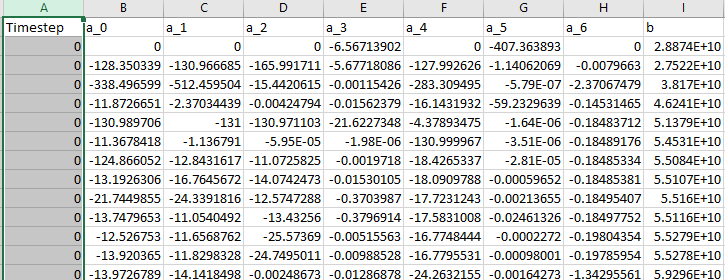


Figure 16: example of Bellman values with 7 reservoirs

It corresponds to the output file BellmanValuesOUT.csv, computed by the SSV. It has the following columns :

* Timestep : index of the SSV timestep ; this timestep is associated to as many rows as there are cuts in the bellmanValues function at this timestep
* N columns corresponding to N réservoirs, named a\_i (i in [0,N-1]) corresponding to the linear terms of the function
* 1 columns named b, corresponding to the fixed term of the function

The simulation tool may also use bellman values which were computed independently, reservoir per reservoir. In this case, one CSV file per reservoir is required, with the following format :



Figure 17: Bellman Values for 1 reservoir

* A first column containing time stamps,
* a second column containing the volumes (in MWh)
* a third column containing the Bellman values.

For a given unit, the number of volume steps must be the same at each time step of the sheet. The function Value(Volume) must be convex. The sheet must contain 1 line per volume ‘step’ per SSVTimeStep.

## Plan4res OUTPUT data

### Outputs of simulation

The simulation outputs a serie of csv files, each one composed of a first row corresponding to the header : names of timeseries, and a first column corresponding to the indexes of timesteps. Eacg CSV file contains one column per timeserie.

The list of files is the following :

* ActiveDemandOUT.csv : 1 column per region ; The names of the series are not (yet) the names of regions (this will be updated in a later version). Each column is named Node\_i, where Node\_i is the ith row of ZP\_ZonePartition.
* ActivePowerOUT.csv : Energy generated at each time step : 1 column per generation unit (2 per generation unit with pumping, ie from ShortTermStorage : 1 for generation and one for consumption used for pumping – negative values- (following the names in columns ‘Name » of sheets TU\_ThermalUnits, STS\_ShortTermStorage, SS\_SeasonalStorage, RES\_RenewableUnits.
* PrimaryOUT.csv : Primary reserve at each timestep: 1 column per generation unit (following the names in columns ‘Name » of sheets TU\_ThermalUnits, STS\_ShortTermStorage, SS\_SeasonalStorage, RES\_RenewableUnits. Contains 0 if the PrimaryReserve constraint is not activated.
* SecondaryOUT.csv : Secondary reserve at each timestep: 1 column per generation unit (following the names in columns ‘Name » of sheets TU\_ThermalUnits, STS\_ShortTermStorage, SS\_SeasonalStorage, RES\_RenewableUnits. Contains 0 if the SecondaryReserve constraint is not activated.
* VolumeOUT.csv : Volumes of all réservoirs at each timestep (includes seasonal storage an short term storage reservoirs) : 2 columns per SeasonalStorage unit (upstream and downstream reservoir), 1 column per ShortTermStorage unit.
* FlowsOUT.csv : flows between regions at each timestep : 1 column per line, named Line\_i where Line\_i corresponds to the ith row of IN\_Interconnection.
* MarginalCostActivePowerDemandOUT.csv : Marginal costs of the ActivePowerDemand constraint at each time step ; 1 column per region.
* MarginalCostPrimaryOUT.csv : Marginal costs of the Primary reservr constraint at each time step ; 1 column per region.
* MarginalCostSecondaryOUT.csv : Marginal costs of the Secondary reserve constraint at each time step ; 1 column per region.
* MarginalCostInertiaOUT.csv : Marginal costs of the inertia constraint at each time step ; 1 column per region.
* MarginalCostFlowsOUT.csv : Marginal costs of the Flows bounds constraints at each time step ; 1 column per line.

### Outputs of Seasonal Storage valuation

* If the SDDP has converged, the outputs (the bellman values) can be found in the file BellmanValuesOUT.csv. The file BellmanValuesAllOUT.csv is also created ; it contains all the cuts that were computed during the optimisation, including useless cuts.
* If the SDDP did not converge, all the cuts that were computed can be found in the file cuts.txt

The format of those 3 files is the following :

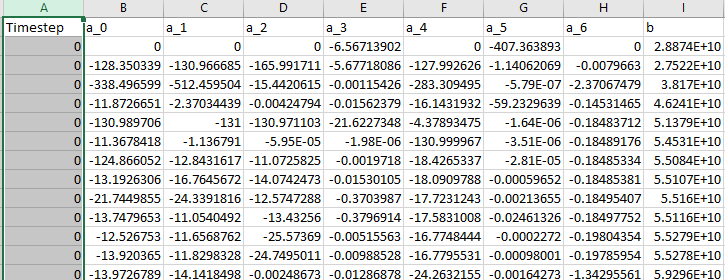


Figure 18: example of Bellman values with 7 reservoirs

This file has the following columns :

* Timestep : index of the SSV timestep ; this timestep is associated to as many rows as there are cuts in the bellmanValues function at this timestep
* N columns corresponding to N réservoirs, named a\_i (i in [0,N-1]) corresponding to the linear terms of the function
* 1 columns named b, corresponding to the fixed term of the function

### Outputs of Capacity Expansion

The CEM creates the file investment\_candidates.txt, which contains all the different investment configurations which were evaluated within the Benders algorithm.

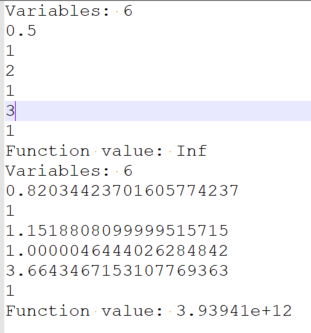


Figure 19: investment\_candidates.txt

The variables correspond to the units for which investment is possible, ie with a MaxAddedCapacity >0 or a MaxRetCapacity<0. The order of the unit sis the following : TU\_ThermalUnits, RES\_RenewableUnits, STS\_ShortTermStorage, IN\_Interconnection. Variables are indexed from 0 to N in the order of each sheet, starting with variable 0 for the first unit with investment capacity in sheet TU\_ThermalUnit, index N-1 corresponds to the last line with expansion capacity in sheet IN\_Interconnection.

For each iteration, the file gives the number of variable, the value of each variable and the value of the cost function. The value of a variable is the multiplying factor to apply to the MaxPower of the unit to obtain the new configuration.

The solution is given by the configuration with the lowest cost. Cost=Inf means that the configuration is not feasible.

The CEM may also run the simulation on all the diffferent scenarios with the optimal investment configuration. It will output the same files as the simulation, except that each file will appear S times if there are S scenarios. Names of the output files will be Name\_Sceni\_OUT.csv (eg Flows\_Scen5\_OUT.csv)