Selecting and Initializing Representative Days for Generation and Transmission Expansion Planning with High Shares of Renewables

G. Micheli[*], M.T. Vespucci[*],
M. Stabile[#], A. Cortazzi[#]

[*] University of Bergamo
[#] CESI, Milano

Outline

- Generation and Transmission Expansion Planning
  - Problem description
- Analysis Performed
  - Selection of representative days
  - Initialization of representative days
- Tests
  - Testing Framework
- Future Work

Generation & Transmission

Expansion Planning

Generation and Transmission Expansion Problem

- Determine a **least-cost** investment schedule for:
    - Construction of new generating capacity;
    - Building of new electrical interconnections;
    - Decommissioning of generating units.

- Given **projections** of the energy system evolution, define expansion plans in order to:
    - Supply load
    - Achieve policy goals

at **minimum** total cost (investment and operation).

Problem Description

- **Long-term** planning horizon;

- **Hourly** resolution needed to catch
  - Fluctuation of solar and wind power generation
  - Technical constraints on thermal power production
    - ✓ Minimum Up Time
    - ✓ Minimum Down Time
  - Dynamics of storage facilities
    - ✓ Hydro pumped storage
    - ✓ Battery storage

- High **uncertainty**.

➡ Computationally **intractable** problem.

Analysis Performed

Selection of Representative Days

- In order to keep the problem computationally tractable frequently **a small number of representative days** is considered.

- Different approaches have been proposed in the **literature**:

  - **Simple heuristics**

    Only days with minimum load, maximum load, largest daily load spread are considered.

  - **Clustering algorithms**

    Days with similar load, wind/solar production are grouped into clusters, with the cluster's centroid then taken as the representative day.

  - **Load duration curve**

    Minimization of the difference between the load duration curve and the one reproduced by the representative days.

Initialization of the Clustering Procedure

- Consider the **first year** of the planning horizon;

- Select days with **minimum** and **maximum** total load in the power system and remove them from the dataset.

- For each day d of the new dataset a vector V_d is created which contains for all zones z the values of
  - ✓ Hourly load $\left(D_{z,t}^{d},\ \ z \in \mathcal{Z},\ 1 \le t \le 24\right)$
  - ✓ Hourly solar capacity factor $\left(\mu_{z,t}^{d},\ \ z \in \mathcal{Z},\ \ 1 \le t \le 24\right)$
  - ✓ Hourly wind capacity factor $\left(\rho_{z,t}^{d},\ \ z \in \mathcal{Z},\ 1 \le t \le 24\right)$

  In this way **correlations** within a day between load and renewable capacity factors in different hours of day and different system zones are taken into account.

- **Normalize** vectors V_d;

- Define a **threshold** for the choice of the number of representative days (e.g. 1%).

Iteration $k = 2, 3, \ldots$

a) By the k-medoids algorithm compute k clusters so as to minimize the deviation between vectors V_d and their representative V_c^*:

$$\min \sum_{c} \sum_{d \in D_c} \|V_d - V_c^*\|^2$$

D_c: group of days d in cluster c

b) Associate to each representative day V_c^* the weight $|D_c|$, i.e., the number of historical days grouped in cluster c

c) Construct the load duration curve corresponding to the representative days and compute its distance to the original load duration curve

d) If the mean absolute percentage error in the load duration curve approximation is lower than the input threshold, stop; otherwise increase k by one and repeat.

Termination of the Clustering Procedure

- The $\boldsymbol{k+2}$ representative days for the first year of the planning horizon have been determined:
  - k centroids identified by the k-medoids algorithm;
  - 2 extreme days.

- Determine the representative days of the subsequent years by applying **annual growth factors** to load profiles.

- Use representative days to evaluate power system **operation** with hourly resolution in the expansion planning model.

Initialization of Representative Days

- The use of representative days raises the crucial issue regarding how these days should be **linked** in the expansion planning model.

- Most of the existing methods consider the representative days as **temporally consecutive**, linking these days according to an arbitrary order
  - The order chosen could affect the model results;
  - The interconnection among days increases computational costs and prevents from exploiting the decomposable structure of the expansion planning problem.

- In our method, we assign to each thermal power plant an **initial** ON/OFF **status** in every representative day by means of a decision tree built on historical data.

Our Proposed Method

- We consider a **training set** including commitment decisions for Italian thermal power plants during a year.

- We build parameters $\gamma_{k\,0}^{d}$, which describe the ON/OFF **status** of thermal power plant k in the last hour of day $d-1$, $2 \leq d \leq 365$.

- We compute the following **features**:
  - Marginal cost ratio;
  - Start-up cost;
  - Minimum up time;
  - Minimum down time.

- We estimate on the training set a **decision tree**, in order to identify a classification rule that could determine the initial ON/OFF status according to features values.

Initialization of Representative Days

- The decision tree is used to assign to each thermal power plant k in every year y of the planning horizon the **probability** π_k^y of having an initial ON status.
  - ✓ Thermal production costs change throughout the planning horizon, thus thermal power plants may present different probabilities π_k^y along the planning horizon.

- Parameters π_k^y are used to **set** the probability of extracting 1 in the random selection between 0 (i.e., OFF) and 1 (i.e., ON).

- For each thermal plant k and for every year y, this random selection is **repeated** for all representative days, in order to assign to each representative day $c \in \mathcal{C}^y$ a specific initial status $\gamma_{k_0}^c$.

Tests

Testing Framework

- A simplified model for GTEP is used as a **testing framework** to assess the performances of the proposed method:
  - System cost minimization formulation;
  - Only thermal, wind and solar power technologies are considered;
  - Zonal representation of power system;
  - Transportation model for power exchanges among zones;
  - Inelastic demand.

Decision Variables: Strategic Decisions

- For each **year** y of the planning horizon
  - $\delta_{k,y}$ Building of new thermal power plant k
  - $\theta_{k,y}$ Availability of new thermal power plant k : $\theta_{k,y} = \sum_{i=1}^{y} \delta_{k,i}$
  - $\delta_{l,y}$ Building of new transmission line l
  - $\theta_{l,y}$ Availability of new transmission line l : $\theta_{l,y} = \sum_{i=1}^{y} \delta_{l,i}$
  - $sol_{z,y}$ New solar installed capacity in zone z
  - $wind_{z,y}$ New wind installed capacity in zone z

Decision Variables: Operational Decisions

- For each **hour** t of every **representative day** c

$p_{k,t}^{c}$	Power output unit k above the minimum
$\gamma_{k,t}^{c}$	1: unit k is ON; 0: otherwise
$\alpha_{k,t}^{c}$	1: unit k is started-up; 0: otherwise
$\beta_{k,t}^{c}$	1: unit k is shut down; 0: otherwise
$x_{l,t}^{c}$	Energy flow on transmission line l
$RES_{z,t}^{c}$	Renewable generation in zone z
$ENP_{z,t}^{c}$	Energy not provided in zone z
$OG_{z,t}^{c}$	Overgeneration in zone z

Objective Function

$$\min z = \sum_{y \in \mathcal{Y}} \frac{1}{(1+r)^{y-y_0}} \left[\sum_{k \in \mathcal{K}_C} IC_k^{th} \ \delta_{k,y} \right.$$

Investment cost for new thermal power plants

$$+ \sum_{z \in \mathcal{Z}} \left(IC_{z,y}^{sol} sol_{z,y} + IC_{z,y}^{wind} wind_{z,y} \right)$$

Investment cost for new RES capacity

$$\left. + \sum_{l \in \mathcal{L}_C} IC_l^{line} \ \delta_{l,y} \right]$$

Investment cost for new transmission lines

$$+ \sum_{y \in \mathcal{Y}} \sum_{c \in \mathcal{C}^y} w_c \sum_{t=1}^{24} \left[\sum_{k \in \mathcal{K}} CM_{k,y} \left(\underline{P}_k \gamma_{k,t}^c + p_{k,t}^c \right) \right.$$

Thermal production cost

$$+ \sum_{k \in \mathcal{K}} SUC_k \alpha_{k,t}^c$$

Start-up cost

$$\left. + c_{ENP} \sum_{z \in \mathcal{Z}} ENP_{z,t}^c + c_{OG} \sum_{z \in \mathcal{Z}} OG_{z,t}^c \right]$$

Penalties

Constraints

- Load Supply
- Reserve Requirements
- Renewables Penetration
- Energy flows on transmission lines
- Operation of thermal units
  - Minimum and maximum power output
  - Minimum up time
  - Minimum down time
  - Consistency between binary variables
    - $\gamma^{c}_{k,t} - \gamma^{c}_{k,t-1} = \alpha^{c}_{k,t} - \beta^{c}_{k,t} \qquad 2 \le t \le 24$
    - $\gamma^{c}_{k,t} - \gamma^{c}_{k_0} = \alpha^{c}_{k,t} - \beta^{c}_{k,t} \qquad t = 1$

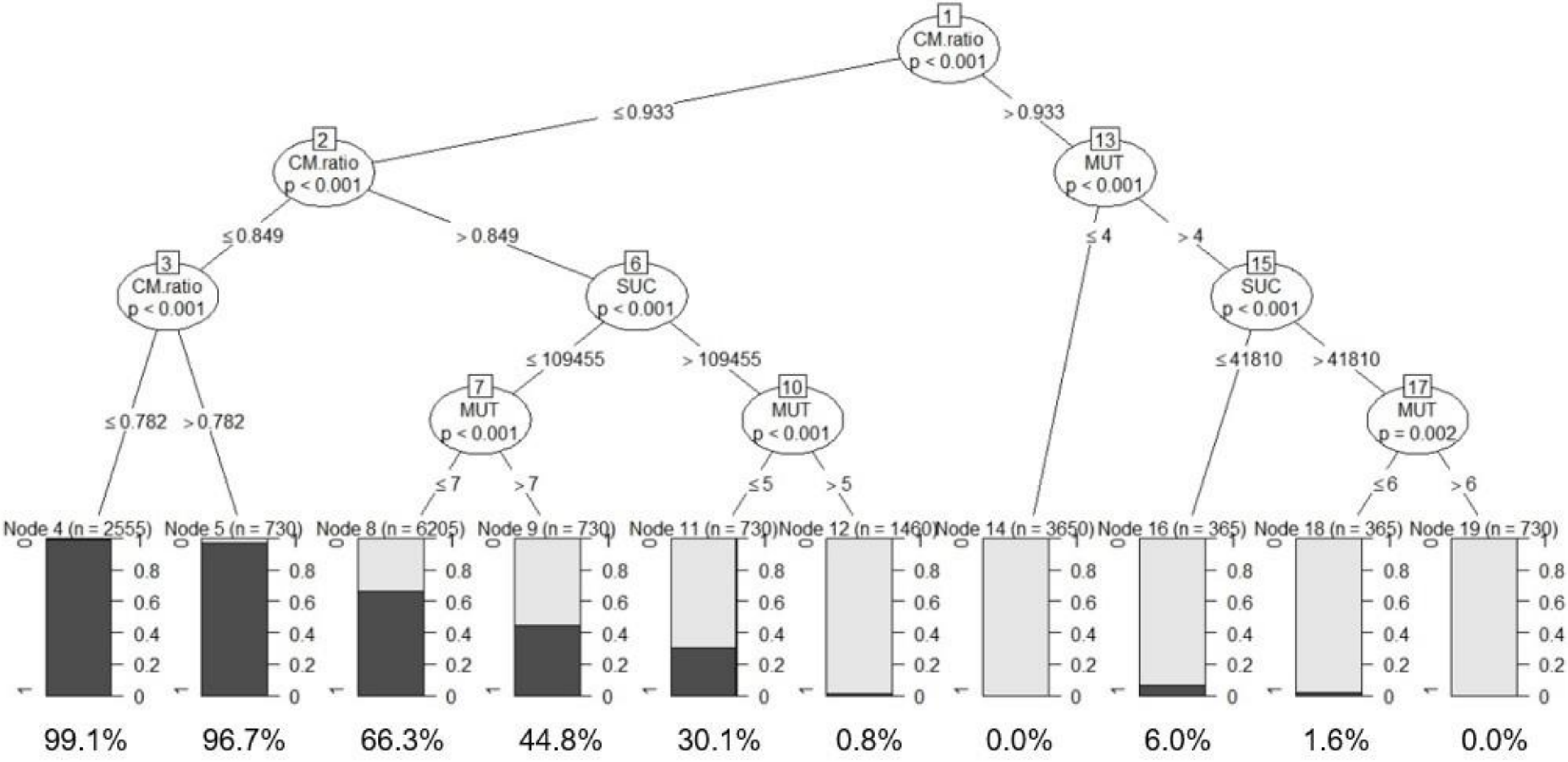
Scenario Data

- Scenario
  - ✓ South-Italy power system
  - ✓ Single year planning horizon
  - ✓ 30% level for RES penetration

- Representative days
  - ✓ Threshold of 1%
  - ✓ 7 representative days

- Initial statuses
  - ✓ Italian power plants
  - ✓ Commitment decisions in 2018

Decision Tree

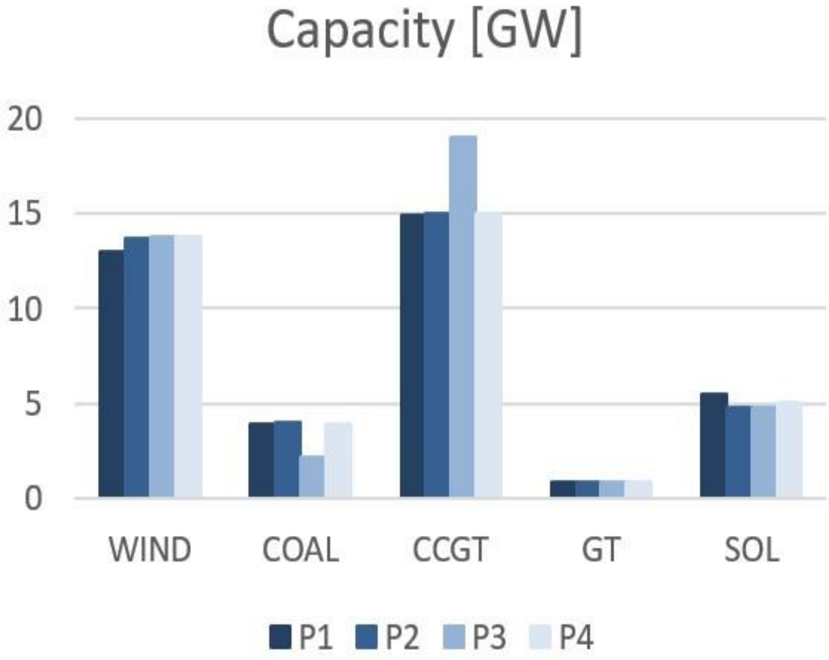


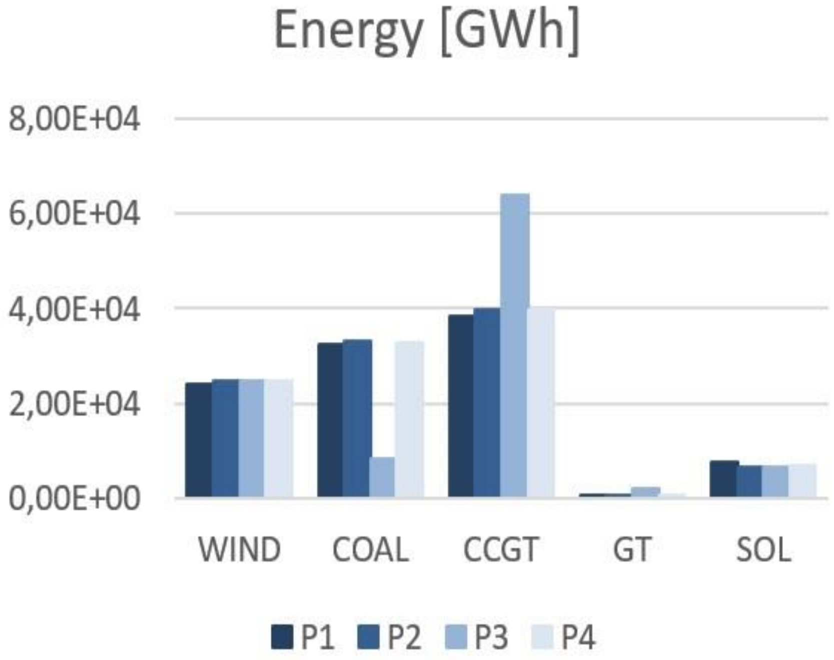
Analysis Validation

- Four formulations are compared:
  - ✓ *P1* Hourly model (i.e., 8760 values for load, solar and wind power capacity factors are considered)
  - ✓ *P2* Representative days are linked, with the initial status of thermal power plants in representative day c being equal to the final status in representative day $c-1$ (i.e., $\gamma_{k0}^{c} = \gamma_{k,24}^{c-1}$)
  - ✓ *P3* Representative days are not linked, but thermal plants are considered offline at the beginning of each day (i.e., $\gamma_{k0}^{c} = 0$)
  - ✓ *P4* Representative days are not linked and parameters γ_{k0}^{c} are determined with the proposed method

CESI

Expansion Plans Comparison: Generation Expansion


Capacity [GW]


Energy [GWh]

Expansion Plans Comparison: Costs and Solution Time

Model	Investment Cost [€]	Production Cost [€]	Start-Up Cost [€]	Total Cost [€]	Total Error	Solution Time [min]
P1	$2.12 \cdot 10^9$	$3.04 \cdot 10^9$	$4.45 \cdot 10^7$	$5.21 \cdot 10^9$	–	393.10
P2	$2.19 \cdot 10^9$	$3.11 \cdot 10^9$	$3.46 \cdot 10^7$	$5.33 \cdot 10^9$	2.50%	3.07
P3	$2.21 \cdot 10^9$	$3.39 \cdot 10^9$	$44.2 \cdot 10^7$	$6.04 \cdot 10^9$	16.09%	2.63
P4	$2.16 \cdot 10^9$	$3.09 \cdot 10^9$	$4.00 \cdot 10^7$	$5.29 \cdot 10^9$	1.64%	2.57

✓ Formulation *P4* presents both the **highest accuracy** and the **lowest solution time**.

✓ Formulation *P4* does not prevent from exploiting the **decomposable** structure of the expansion planning problem given by the use of disconnected representative days.

Future Work

Future Work

- Integration of the proposed procedure in a detailed model for **GTEP**:
  - Thermal, hydro, wind and solar generation
  - Transmission network
  - Energy storage systems
  - Demand side management devices
  - Gas network
  - Power-to-gas facilities
  - Uncertainty inclusion

Thanks for your attention