Updated: 3/5/2024

1 Introduction

In this set of notes, we develop the generation expansion planning (GEP) problem by starting at the simplest version and gradually adding modeling fidelity (at computational expense) until we obtain the most mature problem. The steps we will take are as follows:

Section 2: GEP problem development

- 2.1 Simple GEP statement: single period, single area
- 2.2 Reserve constraint
- 2.3 Capacity constraint
- 2.4 Accounting for different demands
- 2.5 Accounting for annual energy production
- 2.6 Representing multiple areas
- 2.7 Representing transmission, but without limits
- 2.8 Representing transmission, with limits
- 2.9 Including adequacy constraints
- 2.10 Multiperiod formulation
- 2.11 End effects

Section 3: Overview of commercial tools

2 GEP problem development

2.1 Simple GEP statement: single period, single area

A simple statement of a single-period, single area generation expansion planning (GEP) problem is as follows:

Problem GEP-1:

$$\min \underbrace{\sum_j I_j Cap_j^{add}}_{Investment\,Costs} + \underbrace{\sum_j FC_j H_j P_j T}_{Operational\,Costs} \qquad (1)$$

subject to

$$Cap_j = Cap_j^{existing} + Cap_j^{add} \qquad \forall j \qquad (2)$$

$$\sum_j P_j = d \qquad (3)$$

$$0 \le P_j \le Cap_j \qquad \forall j \qquad (4)$$

$$Cap_j^{add} \ge 0 \qquad \forall j \qquad (5)$$

Variable definitions follow:

- j: index over technologies
- I_j: investment cost for technology j, \$/MW
- Cap_j^{add}: capacity added for technology j, MW
- FC_j: fuel cost for technology j, \$/MBTU
- H_j: heat-rate for technology j, MBTU/MWhr
- P_j: Power generation level for technology j, MW
- T: Planning horizon, hours

In GEP-1, (1) is the objective function which minimizes the investment costs plus the operational costs. There is no need to account for the time value of money here since this is only a single period formulation.

- Equation (2) computes total capacity from existing and added capacity, for each technology *j*.
- Equation (3) expresses power balance for just a single demand level.
- Equation (4) requires power generation level to be within unit capacity.
- Equation (5) imposes non-negativity on the C_j^{add} variables.

The decision-variables for this problem are Cap_j^{add} and P_j.

2.2 Reserve constraint

Problem GEP-1 imposes only that capacity be sufficient to meet the demand, via (3) and (4). However, we also need reserves. Therefore we include a reserve constraint via (6), according to Problem GEP-2:

Problem GEP-2:

$$\min \underbrace{\sum_j I_j Cap_j^{add}}_{InvestmentCosts} + \underbrace{\sum_j FC_j H_j P_j T}_{OperationalCosts} \tag{1}$$

subject to

$$Cap_j = Cap_j^{existing} + Cap_j^{add} \quad \forall j \tag{2}$$

$$\sum_j P_j = d \tag{3}$$

$$0 \le P_j \le Cap_j \quad \forall j \tag{4}$$

$$Cap_j^{add} \ge 0 \quad \forall j \tag{5}$$

$$\sum_j Cap_j \ge (1+r)d \tag{6}$$

Here, r is planning reserve margin (PRM), the fraction required of capacity over peak demand. Most N. American regions require a PRM of between 10 and 20%. Figure 2-1 [1] captures this for all N. American regions in the year 2020; Figure 2-2 [2] captures this for MISO in the year 2023. In these figures, the anticipated includes firm new builds and retirements; the prospective includes firm and likely new builds and retirements; and the reference indicates a target level of planning reserve margin.

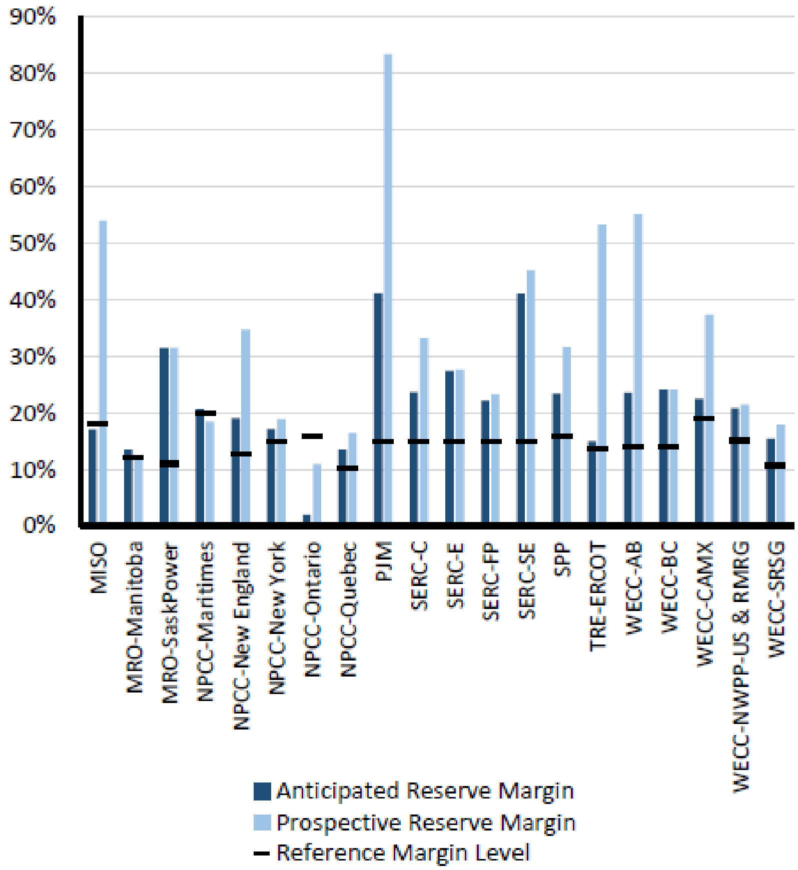


Figure 2-1: Planning reserve margins in N. America, 2020 [1]

Figure 2-2: Planning reserve margins in MISO, 2023 []

2.3 Capacity credit

As we have seen in previous EE 552 notes, the capacity credit of a resource is used to identify the percentage of the resource's capacity which should be identified for reliability calculations at peak load. The capacity credit indicates the availability at peak load.

The capacity credit for wind is usually fairly low because wind generation during daytime hours tends lower. For example, MISO was using 13.3% capacity credit for wind in 2014 [3]; it increased to 16.3% for 2021 [4]; it is 18.1% in summer 2023, 23.1% in fall 2023, 40.3% in winter 2024, and 23.0% in spring 2024 [5].

Capacity credits for solar can be higher because peak loads usually occur during the daytime when solar irradiance (watts/m^2)[1] is highest. MISO used 50% for their 2021 planning cycle [4], and in 2023-2024, maintained that with the exception of winter 2024 when they reduced it to 5% (due to the fact that winter days are so much shorter so that the sun is almost down during the late afternoon/early evening time of typical peaks). Computing capacity credit for wind and solar is an interesting problem that involves calculation of the expected load carrying capacity (ELCC). We reviewed this concept earlier in the course.

We modify (6) to account for capacity credit, CC_j, for each technology j.

[1] Irradiance is a measure of power density, in watts/m^2. Another related measure is insolation, which is a measure of energy density per unit time, e.g., watt-hours/m^2/year.

Problem GEP-3:

$$\min \underbrace{\sum_j I_j Cap_j^{add}}_{Investment Costs} + \underbrace{\sum_j FC_j H_j P_j T}_{Operational Costs} \qquad (1)$$

subject to

$$Cap_j = Cap_j^{existing} + Cap_j^{add} \quad \forall j \qquad (2)$$

$$\sum_j P_j = d \qquad (3)$$

$$0 \le P_j \le Cap_j \quad \forall j \qquad (4)$$

$$Cap_j^{add} \ge 0 \quad \forall j \qquad (5)$$

$$\sum_j CC_j Cap_j \ge (1+r)d \qquad (6a)$$

Observe we have only applied CC_j to (6a), but not to (4). We will address this issue at the end of the next section.

2.4 Accounting for different demands

The GEP statements defined so far require one to choose a particular demand d. In this case, there are problems encountered, as follows:

- If d is chosen to be the peak load, then the GEP will build the right amount of capacity but will over-estimate the energy requirements and corresponding generation production.
- If d is chosen to be the average demand, then the GEP will identify the right energy requirements and corresponding generation production but will underestimate the capacity.

In both cases, the solution will identify the one technology that minimizes the sum of investment plus fuel costs.

An improvement on this formulation is to increase the number of load levels, or load *blocks*, so that we use d_s, s=1… instead of just d. In doing this, we must also identify the appropriate duration h_s for each load block.

A tool to use in identifying appropriate load blocks and corresponding durations is the load duration curve (LDC). We have encountered LDC previously in EE 552 discussions; we repeat the development of LDCs in the Appendix. Figure 2-3 illustrates an LDC, where we observe that it provides on the abscissa the number of hours the load is expected to be greater than or equal to the corresponding load given as the ordinate.

In Figure 2-3, for example, we observe that the load is greater than or equal to 9 MW for 2628 hrs/year, 7 MW for 6132 hrs/year, and 5 MW for 8760 hrs/year (i.e., it is always greater than or equal to 5 MW).

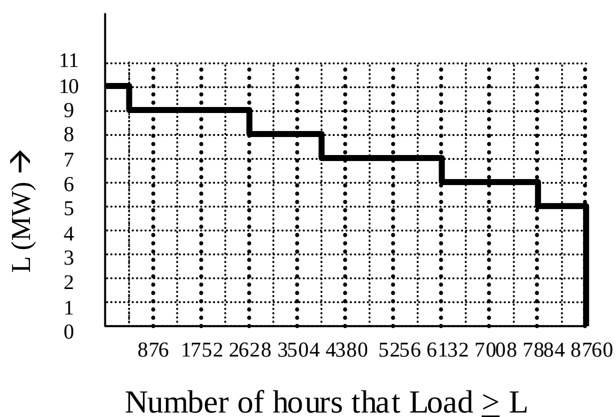


Figure 2-3: Illustration of LDC

We can understand the LDC from a different perspective if we first normalize the abscissa of Figure 2-3 by dividing it by 8760, as shown in Figure 2-4.

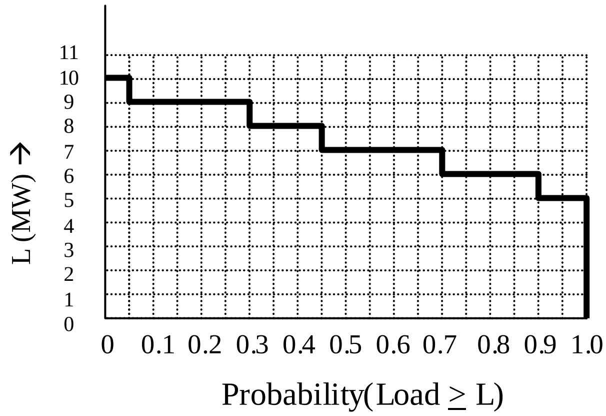


Figure 2-4: LDC with normalized abscissa

Now flip the axes of the plot in Figure 2-4 to show the plot in Figure 2-5. This plot is a cumulative distribution function (CDF) which provides probabilities that the load will be greater than or equal to the value of the abscissa.

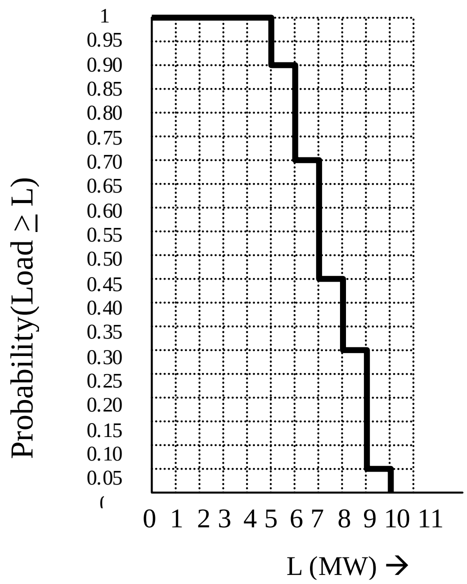


Figure 2-5: Cumulative Distribution Function

For example, we observe from Figure 2-5 that

$$\Pr(\text{Load} \geq 7) \approx 0.45$$
$$\Pr(\text{Load} \geq 8) \approx 0.30$$

We find from the appendix (where exact calculations are illustrated) that the actual values are

$$\Pr(\text{Load} \geq 7) = 0.478$$
$$\Pr(\text{Load} \geq 8) = 0.304$$

We now show how to utilize the LDC in our GEP formulation. Consider the notation associated with the LDC as shown in Figure 2-6. Only 3 load blocks are shown, but the formulation generalizes to any desired number of load blocks. (Formulations GEP-1, GEP-2, and GEP-3 can be viewed as special cases where there was only one load block).

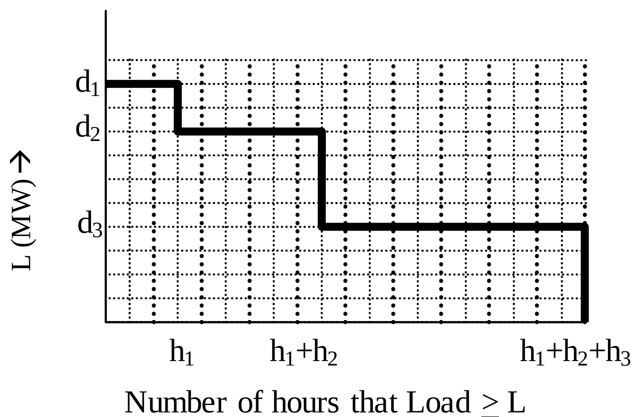


Figure 2-6: LDC with notation for GEP formulation

The GEP, modified to account for the different number of load blocks, is provided below.

Problem GEP-4:

$$\min \underbrace{\sum_j I_j Cap_j^{add}}_{Investment Costs} + \underbrace{\sum_j FC_j H_j \sum_s P_{j,s} h_s}_{Operational Costs} \qquad (1a)$$

subject to

$$Cap_j = Cap_j^{existing} + Cap_j^{add} \qquad \forall j \qquad (2)$$

$$\sum_j P_{j,s} = d_s \qquad \forall s \tag{3a}$$

$$0 \le P_{j,s} \le Cap_j \qquad \forall j, s \tag{4a}$$

$$Cap_j^{add} \ge 0 \qquad \forall j \tag{5}$$

$$\sum_j CC_j Cap_j \ge (1+r) d_1 \tag{6b}$$

Observe that the following changes were made:

- The energy term P_jT in (1) was modified in (1a) to become the summation over s of $P_{j,s}h_s$.
- The summation over j of the operational values P_j in (3), required to equal d, was modified in (3a) so that such a summation occurs for each load block s.
- The constraints on each operational value P_j to not exceed capacity in (4) was modified in (4a) so that these constraints are imposed for each load block s.
- The load d in the reserve constraint (6a) (which only dealt with a single load block) is specified in (6b) to be the peak load d_1.

Only (2) and (5) remain the same.

There is another issue that deserves some attention. We generally know when the various load levels s=1,2,3 occur. For example, the peak condition may occur in early evening, the off-peak condition in the middle of the night, and, a partial peak condition close to noon, as illustrated for a MISO March day in Figure 2-7. The times at which these various load levels occur can also be seasonal, e.g., for a MISO colder day, the peak time may occur earlier in the morning due to increased electric heating load earlier in the morning, and no air conditioning load at all throughout the day, as shown in Figure 2-8.

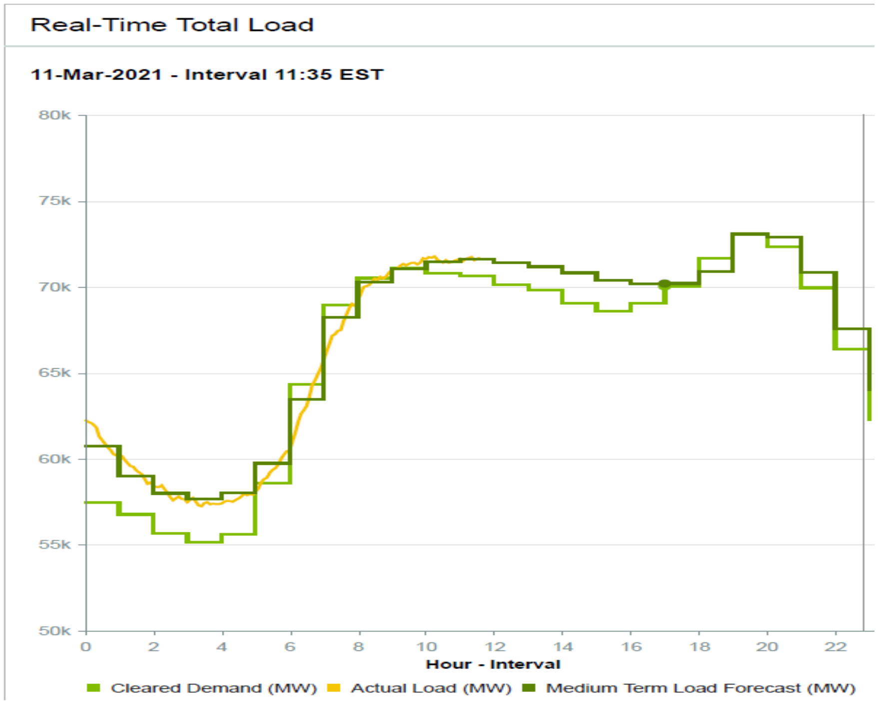


Figure 2-7: MISO real-time load curve, 3/11/2021

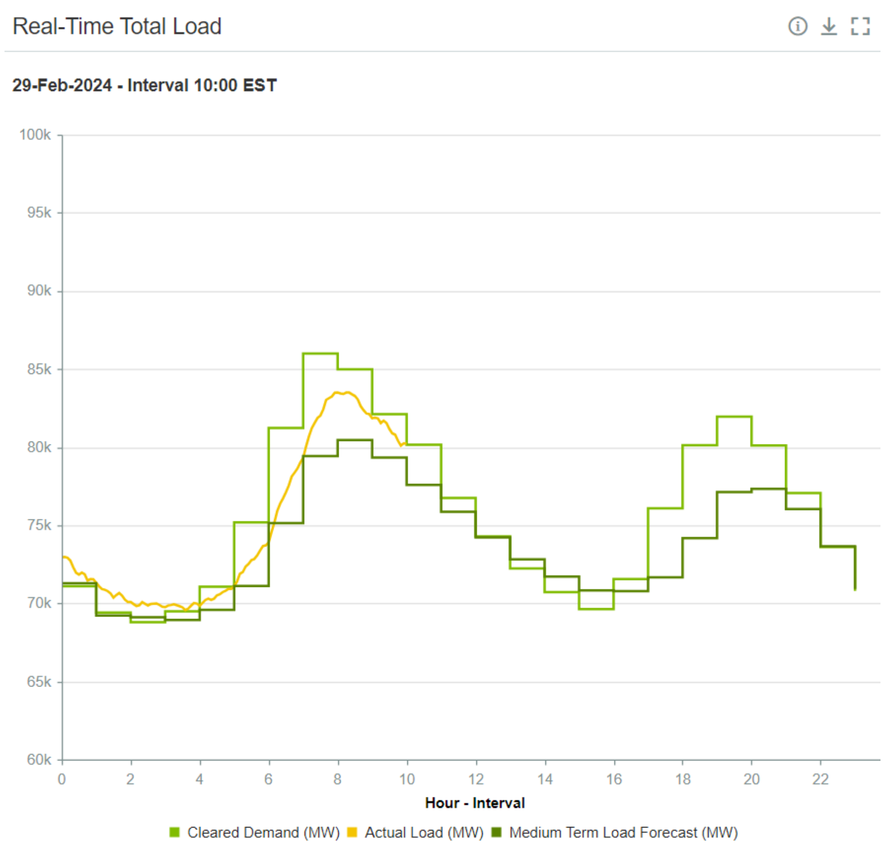


Figure 2-8: MISO real time load curve, 2/29/2024

We need to address two issues here.

First, it is typical to have, for example, 4 load blocks per day (peak, two partial peaks, and an off-peak), and four seasons per year. This makes 16 conditions per year, to which we add load block #1 to account for condition corresponding to the annual peak (d_1). This results in N_s=17 conditions per year.

It would be reasonable to use two subscripts here, one for the load block corresponding to the time-of-day k, and the other corresponding to the season. In this case, we would have d_{ks}, with subscripts k=1,…,4 (time of day), and s=1,…,4 (seasons), and k=1,s=5 (peak). Then we have the following load levels and hours per load level:

Fall:	d_{11}, d_{21}, d_{31}, d_{41},	h_{11}, h_{21}, h_{31}, h_{41},
Winter:	d_{12}, d_{22}, d_{32}, d_{42},	h_{12}, h_{22}, h_{32}, h_{42},
Spring:	d_{13}, d_{23}, d_{33}, d_{43},	h_{13}, h_{23}, h_{33}, h_{43},
Summer:	d_{14}, d_{24}, d_{34}, d_{44},	h_{14}, h_{24}, h_{34}, h_{44},
Peak:	d_{15}	h_{15}=1.

Instead, to maintain notational simplicity, we will assume that, each load block d_s (which is an operating condition) s=1,…,N_s corresponds to a specific season and a specific daily time period, and that we know the number of hours h_s in the year for which that operating condition occurs. Thus, the above blocks are notated as follows:

Fall:	d_1, d_2, d_3, d_4,	h_1, h_2, h_3, h_4,
Winter:	d_5, d_6, d_7, d_8,	h_5, h_6, h_7, h_8,
Spring:	d_9, d_{10}, d_{11}, d_{12},	h_9, h_{10}, h_{11}, h_{12},
Summer:	d_{13}, d_{14}, d_{15}, d_{16},	h_{13}, h_{14}, h_{15}, h_{16},
Peak:	d_{17}	h_{17}=1.

Second, we elect to generalize the notion of capacity credit so that it is chosen for each of these load block conditions. Thus, our capacity credit nomenclature changes from being specific to each technology, CC_j, to being specific to each technology and load block, $CC_{j,s}$. In addition, we are generalizing its use (and perhaps abusing the terminology) because the notion of capacity credit (or accreditation) is generally reserved for resource adequacy assessment, typically done at the annual peak. Here, we are associating a maximum capability (a capacity), for each technology ***and for each load block***. Thus, for each technology, we obtain the following capacity credits values:

Fall: CC_{j1}, CC_{j2}, CC_{j3}, CC_{j4},
Winter: CC_{j5}, CC_{j6}, CC_{j7}, CC_{j8},
Spring: CC_{j9}, CC_{j10}, CC_{j11}, CC_{j12},
Summer: CC_{j13}, CC_{j14}, CC_{j15}, CC_{j16},
Peak: CC_{j17}

It may be wise to reserve the term "capacity credit" for resource adequacy evaluation (and here use it only for CC_{j17}) and then define another term CM_{js} to identify what we think each technology's maximum output will be in that time of day and season. But we elect to stay with CC_{js} in these notes.

We deploy this notion here in order to limit the generation values of wind and solar. If we do not do this, we will tend to obtain unrealistic values of generation levels $P_{j,s}$, e.g., we may obtain overly-high values of wind output during load block s=13 (which corresponds to a summer morning period), or worse yet, significant solar output during load blocks s=1, 5, 9, or 13 (which corresponds to the night-time periods of the four seasons).

We still utilize the capacity credit for the peak period, s=17, and for the reserve constraint.

These changes lead us to Formulation GEP-5.

Formulation GEP-5:

$$\min \underbrace{\sum_j I_j Cap_j^{add}}_{Investment Costs} + \underbrace{\sum_j FC_j H_j \sum_s P_{j,s} h_s}_{Operationd Costs} \qquad (1a)$$

subject to

$$Cap_j = Cap_j^{existing} + Cap_j^{add} \qquad \forall j \qquad (2)$$

$$\sum_j P_{j,s} = d_s \qquad \forall s \qquad (3a)$$

$$0 \le P_{j,s} \le CC_{j,s} Cap_j \qquad \forall j, s \qquad (4b)$$

$$Cap_j^{add} \ge 0 \qquad \forall j \qquad (5)$$

$$\sum_j CC_{j,17} Cap_j \ge (1+r) d_{17} \qquad (6c)$$

Observe that the following changes were made:

- The upper bound Cap_j on the operational variables $P_{j,s}$ in (4a) are multiplied by $CC_{j,s,}$ in (4b).
- The capacity credit CC_j in the reserve constraint (6b) is specified to be the capacity credit corresponding to the peak period, CC_{j17}, in (6c), and here we specify the demand to be d_{17}.

One observes that, in this formulation, the influences on which technology is chosen depend on the investment cost, the fuel cost and heat rate, the maximum generating capability of each technology in each load block, and the number of hours in each load block. Some technologies, like wind and solar, have low investment cost per MW but do not have high capacity credits; although their investment cost per MW tend to be lower than other technologies, their advantage is reduced relative to other technologies if we consider investment cost per unit of credited capacity (still however, their zero fuel cost generally makes wind and solar look very good).

The dependency of capacity credit on load block can result in this formulation seeing different technologies being least cost from one load block to another. This, at least in theory, can result in solutions providing a combination of technologies in order to achieve capacity benefits from those with low investment costs (e.g. combustion turbines) and energy benefits from technologies with low fuel costs (e.g., wind and solar).

2.5 Accounting for annual energy production

The last change made (CC) accounts for the tendency of each technology to operate within certain ranges during certain time periods (where time periods correspond to load blocks).

Another feature which is as yet unaccounted for is the tendency of each technology to produce over a time frame (e.g., a year) a certain fraction of the energy it would produce if it continuously operated at its capacity during that time frame. This feature is typically captured via the capacity factor, denoted here by CF_j.

We account for this tendency in Formulation GEP-6.

Formulation GEP-6:

$$\min \underbrace{\sum_j I_j Cap_j^{add}}_{InvestmentCosts} + \underbrace{\sum_j FC_j H_j \sum_s P_{j,s} h_s}_{OperationalCosts} \qquad (1a)$$

subject to

$$Cap_j = Cap_j^{existing} + Cap_j^{add} \qquad \forall j \qquad (2)$$

$$\sum_j P_{j,s} = d_s \qquad \forall s \qquad (3a)$$

$$0 \le P_{j,s} \le CC_{j,s} Cap_j \qquad \forall j, s \qquad (4b)$$

$$Cap_j^{add} \ge 0 \qquad \forall j \qquad (5)$$

$$\sum_j CC_{j,1} Cap_j \ge (1+r) d_1 \qquad (6c)$$

$$\sum_s P_{j,s} h_s \le CF_j Cap_j \sum_s h_s \qquad \forall j \qquad (7)$$

The only change made to GEP-6, relative to GEP5, is that (7) has been added to the formulation.

It may also be of interest to specify a unique CF for each load block, CF_{js}. In this case, (7) could be replaced by (7') or by (7''):

$$\sum_s P_{j,s} h_s \le Cap_j \sum_s CF_{js} h_{js} \qquad \forall j \qquad (7')$$

$$P_{j,s} h_{js} \le CF_{js} Cap_j h_{js} \qquad \forall j,s \qquad (7'')$$

The difference between (7), (7') and (7'') is that

- (7) requires the same technology CF for each load block; it enforces energy limitations across the entire year;
- (7') allows technology CF for each load block but still enforces energy limitations across the entire year;
- (7'') allows technology CF for each load block and enforces energy limitations for each load block, but it is more computational since it will have N_s times as many constraints.

We will use (7) in what follows.

2.6 Representing multiple areas

So far, we have formulated the GEP for only a single area. Doing so is typical when a single company is performing generation expansion, particularly if the company is not transmission limited. However, when solving GEP under transmission limitations, it is important to represent the GEP in terms of multiple areas.

This means we will also be interested to represent transmission. A preliminary step to representing transmission is to first state the multiarea GEP without transmission. To do this, we introduce the subscript i to denote a particular area, in Formulation GEP-7a.

Formulation GEP-7a:

$$\min \underbrace{\sum_i \sum_j I_{i,j} Cap_{i,j}^{add}}_{InvestmentCosts} + \underbrace{\sum_i \sum_j FC_{i,j} H_j \sum_s P_{i,j,s} h_s}_{OperationalCosts} \qquad (1b)$$

subject to

$$Cap_{i,j} = Cap_{i,j}^{existing} + Cap_{i,j}^{add} \quad \forall i, j \qquad (2a)$$

$$\sum_i \sum_j P_{i,j,s} = \sum_i d_{i,s} \quad \forall s \qquad (3b)$$

$$0 \le P_{i,j,s} \le CC_{i,j,s} Cap_{i,j} \quad \forall i, j, s \qquad (4c)$$

$$Cap_{i,j}^{add} \ge 0 \quad \forall i, j \qquad (5a)$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j} \ge (1+r) \sum_i d_{i,17} \qquad (6d)$$

$$\sum_s P_{i,j,s} h_s \le CF_{i,j} Cap_{i,j} \sum_s h_s \quad \forall i, j \qquad (7a)$$

Observe that the following changes were made:

- The subscript i was added to all variables and parameters, with the only exceptions being

- r: the reserve requirement, which is considered to be imposed across the entire system;
- h_s: the duration for each load block.

- Investment costs and operational costs in (1b) were summed across all areas i.
- Equation (2a) is written for each technology j, as in (2), and also for each area i.
- Equation (3b) is summed across all areas i on both the left-hand-side (to obtain total generation for load block s) and on the right-hand-side (to obtain the total load for load block s).
- Equation (4c) allows the capacity credit for each technology j and load block s to also be specified for each area i.
- Equation (5a) specifies a $C_{i,j}^{add}$ variable for each area i.
- Equation (6d) sums across all areas i on both the left-hand-side (to obtain the total capacity of the interconnection) as well as on the right-hand-side (to obtain the interconnection's reserve requirement).
- Equation (7a) allows the technology capacity factor constraint (to limit total energy over the time period) to be area-specific.

The implication of (3b) and (6d) are that transmission capacity is unlimited, i.e.,

- (3b) says the total interconnection load must be supplied by the total interconnection generation;
- (6d) says the reserve requirement is system-wide (and not area-specific).

Formulation GEP-7a differs from solving GEP-6 for the entire interconnection in that in GEP-7a, parameters (CF and CC) are geographically differentiated, whereas in GEP-6, there is only one set of parameters.

An alternative formulation would assume that there is no transmission between the areas, as in Formulation 7b.

Formulation GEP-7b:

$$\min \underbrace{\sum_i \sum_j I_{i,j} Cap_{i,j}^{add}}_{InvestmentCosts} + \underbrace{\sum_i \sum_j FC_{i,j} H_j \sum_s P_{i,j,s} h_s}_{OperationalCosts} \quad (1b)$$

subject to

$$Cap_{i,j} = Cap_{i,j}^{existing} + Cap_{i,j}^{add} \quad \forall i,j \quad (2a)$$

$$\sum_j P_{i,j,s} = d_{i,s} \quad \forall i,s \quad (3c)$$

$$0 \le P_{i,j,s} \le CC_{i,j,s} Cap_{i,j} \quad \forall i,j,s \quad (4c)$$

$$Cap_{i,j}^{add} \ge 0 \quad \forall i,j \quad (5a)$$

$$\sum_j CC_{i,j,17} Cap_{i,j} \ge (1+r_i) d_{i,17} \quad \forall i \quad (6e)$$

$$\sum_s P_{i,j,s} h_s \le CF_{i,j} Cap_{i,j} \sum_s h_s \quad \forall i,j \quad (7a)$$

Observe that the following changes were made:

- Equation (3c) is written for each area, thus, each area must satisfy its own demand.
- Equation (6e) is also written for each area, thus, each area must satisfy its own reserve requirement. Notice also that each area's reserve requirement r_i is specific to area i.

Since there is no transmission between the areas, this problem is equivalent to solving GEP-6 for each of the areas.

2.7 Representing transmission, but without limits

The intermediate formulation between Formulations GEP-7a (unlimited transmission) and GEP-7b (no transmission) is to represent capacitated (i.e., limited) transmission. We will do this in the next subsection, but first, we represent transmission without imposing transmission limits. Here, each area i is considered to be a network node.

We extend this from GEP-7a, except that, instead of (3b) (which is an interconnection-wide power balance equation), we use an extended form of GEP-7b's node-specific power balance equation (3c), as it facilitates use of a DC power flow formulation. Formulation GEP-8:

$$\min \underbrace{\sum_i \sum_j I_{i,j} Cap_{i,j}^{add}}_{InvestmentCosts} + \underbrace{\sum_i \sum_j FC_{i,j} H_j \sum_s P_{i,j,s} h_s}_{OperationalCosts} \qquad (1b)$$

Note: all quantities must be in per-unit.

subject to

$$Cap_{i,j} = Cap_{i,j}^{existing} + Cap_{i,j}^{add} \qquad \forall i, j \qquad (2a)$$

$$\sum_j P_{i,j,s} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s} = d_{i,s} \qquad \forall i, s \qquad (3d)$$

$$0 \le P_{i,j,s} \le CC_{i,j,s} Cap_{i,j} \qquad \forall i, j, s \qquad (4c)$$

$$Cap_{i,j}^{add} \ge 0 \qquad \forall i, j \qquad (5a)$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j} \ge (1+r) \sum_i d_{i,17} \qquad (6d)$$

$$\sum_s P_{i,j,s} h_s \le CF_{i,j} Cap_{i,j} \sum_s h_s \qquad \forall i, j \qquad (7a)$$

This formulation is no different Formulation GEP-7a in terms of solution, except that it is more computational (due to its inclusion of the linearized power flow equations) and that it provides flows between areas for each of the load blocks.

The only difference between GEP-8 and GEP-7a is the presence of (3d) instead of (3b). The nomenclature here needs definition, as follows:

- b_l is the negated admittance of branch l (therefore +), without j (or one can say it is the negated susceptance ("b" in y=g-jb)
- $S_{l.i}$ is element (l,i) of the node-arc incidence matrix.
- $S_{l.k}$ is element (l,k) of the node-arc incidence matrix.
- $\theta_{k,s}$ is the angle of node (or area) k under loading block s.

A word about the node-arc incidence matrix is in order. It is also called adjacency matrix, or connection matrix. This matrix is well known in any discipline having reason to structure its problems using a network of nodes and "arcs" (or branches or edges). Any type of transportation engineering is typical of such a discipline.

The node-arc incidence matrix contains a number of rows equal to the number of arcs and a number of columns equal to the number of nodes.

Element (l,i) of A is 1 if the l^{th} branch begins at node i, -1 if the l^{th} branch terminates at node i, and 0 otherwise.

A branch is said to "begin" at node i if power flowing across branch l is defined + for a direction from node i to the other node.

A branch is said to "terminate" at node i if the power flowing across branch l is defined positive for a direction to node l from the other node. Figure 2-9 is an illustration of a node-arc incidence matrix for a particular network.

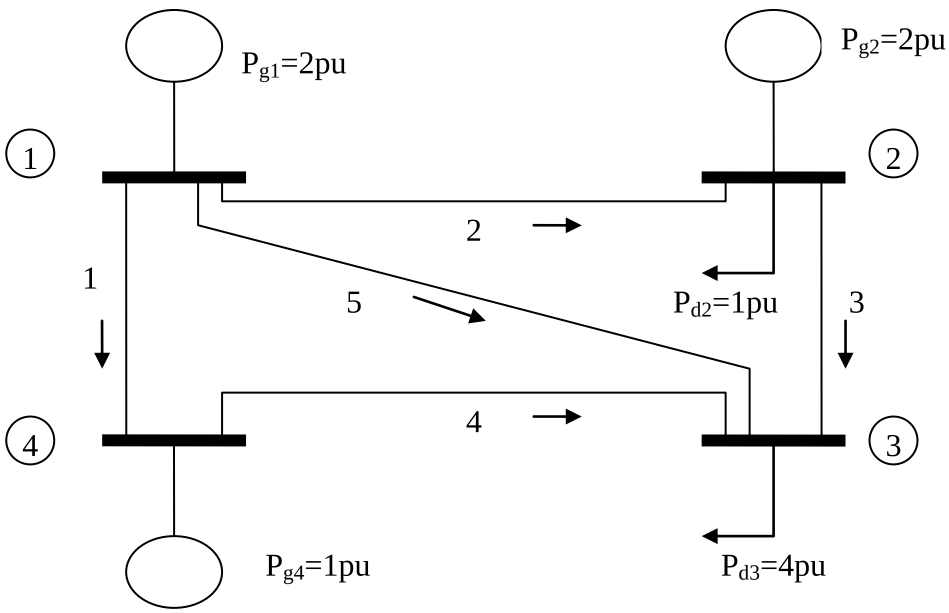


Figure 2-9: Example network for illustrating adjacency matrix

The admittance for all branches is y = g-jb = –j10, (implying it is reactive, not capacitive), so that the susceptance is b=10.

$$\underline{A} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 \end{bmatrix}$$

Reconsider (3d):

$$\sum_j P_{i,j,s} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s} = d_{i,s} \qquad \forall i,s \tag{3d}$$

In this equation:

- b_l is a positive number;
- For each branch l, there will be one product term $S_{l,i}$ $S_{l,k}$ equal to 1 (when k=i) and one product term $S_{l,i}S_{l,k}$ equal to -1 (when

k is a terminating node for a branch connected to node i); all other product terms $S_{l,i}S_{l,k}$ will be zero.

ASIDE: We illustrate the above relation. To simplify, we drop subscripts s (for LDC blocks) and j (for technologies); we have

$$P_i - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_k = d_i$$

We write the expression using the above system for node 1:

$$P_1 - \sum_l \sum_k b_l S_{l,1} S_{l,k} \theta_k = d_1$$

We observe that there are 5 branches, so that l=1,…,5, and there are 4 nodes, so that k=1,…,4. Therefore:

$$P_1 - \sum_{l=1}^{5} \sum_{k=1}^{4} b_l S_{l,1} S_{l,k} \theta_k = d_1$$

Expanding the inner summations results in

$$P_1 - \sum_{l=1}^{5} b_l S_{l,1} \left(S_{l,1}\theta_1 + S_{l,2}\theta_2 + S_{l,3}\theta_3 + S_{l,4}\theta_4 \right) = d_1$$

Now expand the outer summation to obtain:

$$\begin{aligned}
&P_1 - b_1 S_{1,1}\left(S_{1,1}\theta_1 + S_{1,2}\theta_2 + S_{1,3}\theta_3 + S_{1,4}\theta_4\right) \\
&- b_2 S_{2,1}\left(S_{2,1}\theta_1 + S_{2,2}\theta_2 + S_{2,3}\theta_3 + S_{2,4}\theta_4\right) \\
&- b_3 S_{3,1}\left(S_{3,1}\theta_1 + S_{3,2}\theta_2 + S_{3,3}\theta_3 + S_{3,4}\theta_4\right) \\
&- b_4 S_{4,1}\left(S_{4,1}\theta_1 + S_{4,2}\theta_2 + S_{4,3}\theta_3 + S_{4,4}\theta_4\right) \\
&- b_5 S_{5,1}\left(S_{5,1}\theta_1 + S_{5,2}\theta_2 + S_{5,3}\theta_3 + S_{5,4}\theta_4\right) \\
&= d_1
\end{aligned}$$

Recalling the node-arc incidence matrix below, we may fill in the values for $S_{l,k}$.

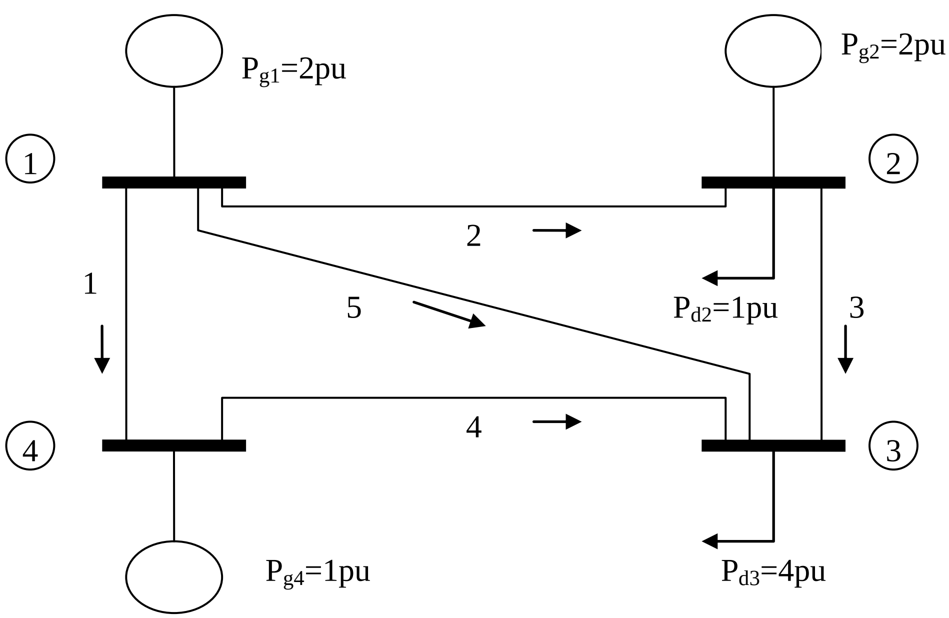
$$\underline{A} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 \end{bmatrix}$$

$$\begin{aligned} &P_1 - b_1(\theta_1 - \theta_4) \\ &- b_2(\theta_1 - \theta_2) \\ &- 0 \\ &- 0 \\ &- b_5(\theta_1 - \theta_3) \\ &= d_1 \end{aligned}$$

This may be re-written as

$$P_1 - d_1 = b_1(\theta_1 - \theta_4) + b_2(\theta_1 - \theta_2) + b_5(\theta_1 - \theta_3)$$

from which it is easy to see the correspondence to the one-line diagram, repeated below for convenience.



2.8 Representing transmission, with limits

Imposing limits on GEP-8 results in Formulation GEP-9:

$$\min \underbrace{\sum_i \sum_j I_{i,j} Cap_{i,j}^{add}}_{InvestmentCosts} + \underbrace{\sum_i \sum_j FC_{i,j} H_j \sum_s P_{i,j,s} h_s}_{OperationalCosts} \qquad (1b)$$

subject to

$$Cap_{i,j} = Cap_{i,j}^{existing} + Cap_{i,j}^{add} \qquad \forall i, j \qquad (2a)$$

$$\sum_j P_{i,j,s} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s} = d_{i,s} \qquad \forall i, s \qquad (3d)$$

$$0 \le P_{i,j,s} \le CC_{i,j,s} Cap_{i,j} \qquad \forall i, j, s \qquad (4c)$$

$$Cap_{i,j}^{add} \ge 0 \qquad \forall i, j \qquad (5a)$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j} \ge (1+r) \sum_i d_{i,17} \qquad (6d)$$

$$\sum_s P_{i,j,s} h_s \le CF_{i,j} Cap_{i,j} \sum_s h_s \qquad \forall i, j \qquad (7a)$$

$$b_l \left| \sum_i S_{l,i} \theta_{i,s} \right| \le F_l^{\max} \qquad \forall l, s \qquad (8)$$

Equation (8) can be expressed as two inequalities, written as

$$-F_l^{\max} \le b_l \sum_i S_{l,i} \theta_{i,s} \le F_l^{\max} \qquad \forall l, s$$

Formulation GEP-9 is exactly like Formulation GEP-8 except for the addition of (8) which limits the flows on the branches. One notes that the linearization necessary for application of DC power flow diminishes in accuracy for angular separations (θ_k-θ_j for line k-j) beyond about 35°. Eq. (8), which results in expressions like -$F_l \le b_l(\theta_k - \theta_j) \le F_l$ for branch l connecting nodes k and j, can serve as a proxy to enforce this constraint. However, this formulation does not allow transmission expansion; it is only modeling limits on the extent that power may be shared between areas.

2.9 Including adequacy constraints

Adequacy evaluation is the evaluation of the extent to which the available generation and transmission is sufficient to supply the load considering steady-state limitations on components (gens & ccts) and their potential for failure. The multi-area reliability analysis we saw in Mod U21 (using convolution or max-flow/decomposition), or what GE-MARS does, is adequacy evaluation.

We augment Formulation GEP-9 with a constraint on an adequacy index. For example, we may constrain loss-of-load probability (LOLP) according to the following:

Formulation GEP-10:

$$\min \underbrace{\sum_i \sum_j I_{i,j} Cap_{i,j}^{add}}_{InvestmentCosts} + \underbrace{\sum_i \sum_j FC_{i,j} H_j \sum_s P_{i,j,s} h_s}_{OperationalCosts} \tag{1b}$$

subject to

$$Cap_{i,j} = Cap_{i,j}^{existing} + Cap_{i,j}^{add} \quad \forall i,j \tag{2a}$$

$$\sum_j P_{i,j,s} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s} = d_{i,s} \quad \forall i,s \tag{3d}$$

$$0 \le P_{i,j,s} \le CC_{i,j,s} Cap_{i,j} \quad \forall i,j,s \tag{4c}$$

$$Cap_{i,j}^{add} \ge 0 \quad \forall i,j \tag{5a}$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j} \ge (1+r) \sum_i d_{i,17} \tag{6d}$$

$$\sum_s P_{i,j,s} h_s \le CF_{i,j} Cap_{i,j} \sum_s h_s \quad \forall i,j \tag{7a}$$

$$b_l \left| \sum_i S_{l,i} \theta_{i,s} \right| \le F_l^{\max} \quad \forall l,s \tag{8}$$

$$LOLP \le LOLP_0 \tag{9}$$

where $LOLP_0$ is a chosen maximum level of LOLP.

The problem with Formulation GEP-10 is that computation of LOLP is intensive, as we have observed in our work with multi-area reliability analysis. There has been work to include adequacy constraints within the GEP problem [6, 7, 8, 9, 10, 11, 12, 13, 14, 15], and it is of interest to study the various approaches to determine the additional computation necessary. One approach that is perhaps most straightforward is to *exclude* the adequacy constraint within the optimization problem and instead evaluate the adequacy of the system after obtaining a solution to the optimization problem; if the adequacy level is not acceptable, then adjustments are made (e.g., by increasing reserve requirements), and the optimization is performed again. Such an approach, illustrated in Figure 2-10, is amendable to optimization methods which employ decomposition (not the same as the decomposition we learned for multiarea reliability analysis). One well-known method is called Benders decomposition; we will study it later.

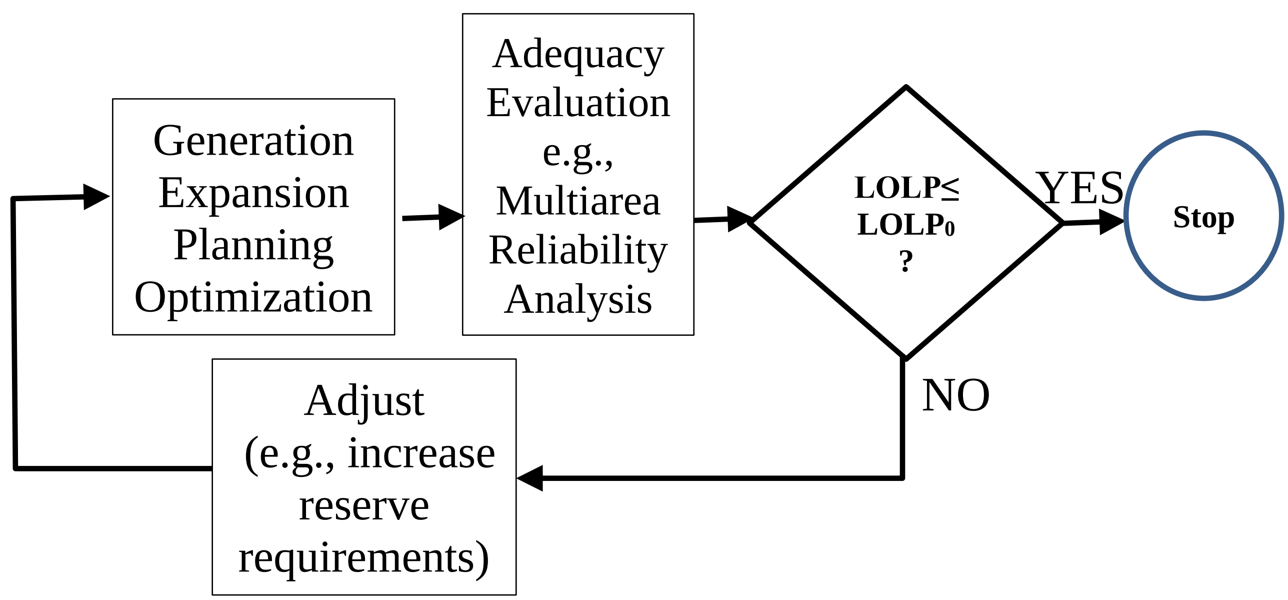


Figure 2-10: GEP with adequacy evaluation

A similar approach is illustrated in Figure 2-11, but here, we use a co-optimized expansion planning (CEP) application which optimizes generation and transmission (step 1), but there are now three functions external to the expansion planning application: multiarea adequacy assessment (steps 2, 3, 4, 5); translation (step 6); and production simulation (i.e., PCM, steps 7, 8, 9). The translation function "translates" the investments from the reduced model to a full ("large") model. The production simulation step is motivated by the need to perform high-fidelity operational modeling (using full model with inter-temporal operational representation instead of the blocks used in the expansion planning function) in order to model flexibility requirements (for regulation reserves, contingency reserves, and ramping reserves).

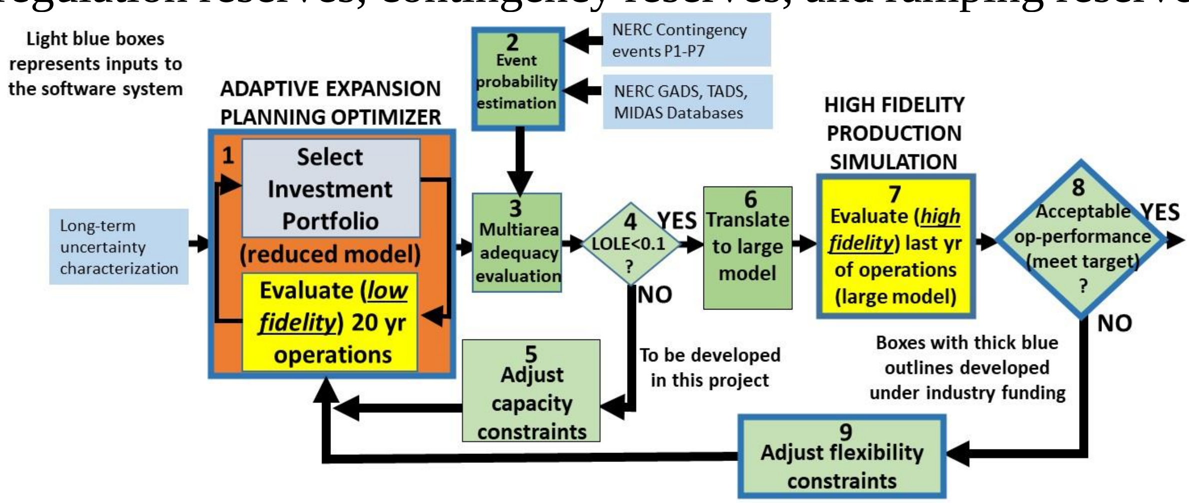


Figure 2-11: Cooptimized expansion planning with adequacy evaluation

2.10 Multiperiod formulation

The problem we described up until this point has retained the single-period simplification. However, a realistic version of the GEP is necessarily multi-period, to account for (i) variation in time when new facilities come on-line; (ii) variation in time when existing facilities are retired; (iii) demand growth, and (iv) the need to meet demand and planning reserve requirements but without overbuilding.

When modeling the single-period GEP, there is no need to handle retirements, because anything to be retired in the period can be reflected in the existing capacity. And so the capacity update equation (2a) works fine.

$$Cap_{i,j} = Cap_{i,j}^{existing} + Cap_{i,j}^{add} \qquad \forall i, j \tag{2a}$$

However, when running a multi-period GEP, retirements occur at different periods, and their effect on each period needs to be reflected in the total available capacity for that year. In fact, retirements, along with demand growth, are the two primary drivers for generation investment.

To this end, we formulate the multi-period GEP with fixed/known retirements, accounting for salvage value.

> Salvage value is the net sum to be realized from the disposal of an asset (net of disposal costs) at the time of its replacement or resale, or at the end of the study period [19, pg. 134].

Formulation GEP-11 accounts for retirements and modeling of salvage values. There are 5 salient changes to our formulation:

- Inclusion of the discount factor ζ in the objective function.
- Provision of subscript t - this subscript t goes from t=1,…T, where T is chosen to reflect the extended time; it affects all equations in the formulation.
- Inclusion of the salvage value in objective function.
- Inclusion of retirement values in the update equation (2b-i).
- Addition of (2b-ii) to initialize capacities to existing values.

Formulation GEP-11:

$$\min \underbrace{\sum_t \zeta^{t-1}\left\{\sum_i \sum_j I_{i,j,t} Cap_{i,j,t}^{add} - SV_{i,j,t} Cap_{i,j,t}^{ret}\right\}}_{\text{Investment Costs Less Salvage Values for Retired Units}}$$

$$+\underbrace{\sum_t \zeta^{t-1}\left\{\sum_i \sum_j FC_{i,j} H_j \sum_s P_{i,j,s,t} h_s\right\}}_{\text{Operational Costs}} \tag{1c}$$

subject to

$$Cap_{i,j,t} = Cap_{i,j,t-1} + Cap_{i,j,t}^{add} - Cap_{i,j,t}^{ret} \qquad \forall i,j,t \tag{2b-i}$$

$$Cap_{i,j,0} = Cap_{i,j}^{exist} \qquad \forall i,j \tag{2b-ii}$$

$$\sum_j P_{i,j,s,t} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s,t} = d_{i,s,t} \qquad \forall i,s,t \tag{3e}$$

$$0 \le P_{i,j,s,t} \le CC_{i,j,s} Cap_{i,j,t} \qquad \forall i,j,s,t \tag{4c}$$

$$Cap_{i,j,t}^{add} \ge 0 \qquad \forall i,j,t \tag{5a}$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j,t}(t) \ge (1+r) \sum_i d_{i,1,t7} \qquad \forall t \tag{6d}$$

$$\sum_s P_{i,j,s,t} h_s \le CF_{i,j} Cap_{i,j,t} \sum_s h_s \qquad \forall i,j,t \tag{7a}$$

$$b_l \left| \sum_i S_{l,i} \theta_{i,s,t} \right| \le F_l^{\max} \qquad \forall l,s,t \tag{8}$$

We describe the modified equations in what follows:

- Equation (1c): There are three changes to describe within the objective function:
    - The discount factor is given by (ζ="zeta"), $\zeta=1/(1+i)$ where i is discount rate. Thus we have that

$$\zeta^t = \frac{1}{(1+i)^t}$$

We assume the investments made in year 1 are already present value, and so it is not until year 2 that we need to discount to present worth; therefore we utilize ζ^{t-1} as the discount factor.

- The investment cost $I_{i,j,t}$, in \$/MW, is subscripted by time t to reflect the fact that it can change over time due to inflation and/or technology maturation.
- The capacity retired for each technology j, in area i, at time t is denoted $Cap^{ret}_{i,j,t}$, and we obtain a salvage value (in \$) by multiplying $Cap^{ret}_{i,j,t}$ (in MW) by $SV_{i,j,t}$ (in \$/MW). A key observation here is that $Cap^{ret}_{i,j,t}$ is a decision variable.
- $SV_{i,j,t}$ (like $I_{i,j,t}$) is given in units of \$/MW and is an input parameter.

* Equation (2b-i):
    - Since this formulation is multi-period, the capacity update equation is recursive, i.e., it updates for each time step.
    - The capacity update equation also includes the effect of retirements.
* Equation (2b-ii): Since the capacity update equation is recursive, we must initialize it to the existing capacity.

An interesting question arises for this formulation. When would capacity ever be retired?

The answer to this question may be obtained by studying the objective function. Recognizing that the optimization is going to choose its decision variables Cap^{add}, Cap^{ret} (and also auxiliary variables P and θ) to minimize the total sum, we can say that it will retire capacity when the savings from salvage value exceeds

- savings obtained via using the unit to displace energy from a more expensive unit over the entire planning horizon (always the case if the unit is never used), AND
- capacity is not needed to satisfy reserves via (6d).

Here are two interesting extensions to this problem. First, we model fuel cost changing with time in the objective, i.e., in GEP-11, we used $FC_{i,j}$ in (1c); here, in GEP-11a, we use $FC_{i,j,t}$ in (1c).

Formulation GEP-11a:

$$\min \underbrace{\sum_t \zeta^{t-1}\left\{\sum_i \sum_j I_{i,j,t} Cap_{i,j,t}^{add} - SV_{i,j,t} Cap_{i,j,t}^{ret}\right\}}_{\text{Investment Costs Less Salvage Values for Retired Units}}$$

$$+\underbrace{\sum_t \zeta^{t-1}\left\{\sum_i \sum_j FC_{i,j,t} H_{i,j} \sum_s P_{i,j,s,t} h_s\right\}}_{\text{Operational Costs}} \qquad (1c)$$

subject to

$$Cap_{i,j,t} = Cap_{i,j,t-1} + Cap_{i,j,t}^{add} - Cap_{i,j,t}^{ret} \qquad \forall i,j,t \qquad (2b\text{-}i)$$

$$Cap_{i,j,0} = Cap_{i,j}^{exist} \qquad \forall i,j \qquad (2b\text{-}ii)$$

$$\sum_j P_{i,j,s,t} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s,t} = d_{i,s,t} \qquad \forall i,s,t \qquad (3e)$$

$$0 \le P_{i,j,s,t} \le CC_{i,j,s} Cap_{i,j,t} \qquad \forall i,j,s,t \qquad (4c)$$

$$Cap_{i,j,t}^{add} \ge 0 \qquad \forall i,j,t \qquad (5a)$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j,t}(t) \ge (1+r) \sum_i d_{i,17,t} \qquad \forall t \qquad (6d)$$

$$\sum_s P_{i,j,s,t} h_s \le CF_{i,j} Cap_{i,j,t} \sum_s h_s \qquad \forall i,j,t \qquad (7a)$$

$$b_l \left| \sum_i S_{l,i} \theta_{i,s,t} \right| \le F_l^{\max} \qquad \forall l,s,t \qquad (8)$$

By modeling fuel cost changing with time, we capture the situation where, if fuel increases too much, it will drive some technologies to never produce and thus be retired.

A second extension is to account for fixed O&M (FOM) and variable O&M (VOM). To address this, we introduce a new formulation that includes FOM and VOM. We could have included these terms in the original formulation, but we elected to do it here to emphasize their impact on retirement decisions.

Formulation GEP-11b:

$$\min \sum_t \zeta^{t-1} \left\{ \underbrace{\sum_i \sum_j I_{i,j,t} Cap_{i,j,t}^{add} - SV_{i,j,t} Cap_{i,j,t}^{ret}}_{\text{Investment Costs Less Salvage Values for Retired Unit}} + \underbrace{\sum_i \sum_j OM_{i,j}^{fixed} Cap_{i,j,t}}_{\text{Fixed O\&M Costs}} \right\}$$
$$+ \sum_t \zeta^{t-1} \left\{ \underbrace{\sum_i \sum_j FC_{i,j,t} H_{i,j} \sum_s P_{i,j,s,t} h_s}_{\text{Fuel Costs}} + \underbrace{\sum_i \sum_j OM_{i,j}^{\text{var}} \sum_s P_{i,j,s,t} h_s}_{\text{Variable O\&M Costs}} \right\} \quad (1c)$$

subject to

$$Cap_{i,j,t} = Cap_{i,j,t-1} + Cap_{i,j,t}^{add} - Cap_{i,j,t}^{ret} \qquad \forall i,j,t \qquad (2b\text{-i})$$

$$Cap_{i,j,0} = Cap_{i,j}^{exist} \qquad \forall i,j \qquad (2b\text{-ii})$$

$$\sum_j P_{i,j,s,t} - \sum_l \sum_k b_l S_{l,i} S_{l,k} \theta_{k,s,t} = d_{i,s,t} \qquad \forall i,s,t \qquad (3e)$$

$$0 \le P_{i,j,s,t} \le CC_{i,j,s} Cap_{i,j,t} \qquad \forall i,j,s,t \qquad (4c)$$

$$Cap_{i,j,t}^{add} \ge 0 \qquad \forall i,j,t \qquad (5a)$$

$$\sum_i \sum_j CC_{i,j,17} Cap_{i,j,t}(t) \ge (1+r) \sum_i d_{i,17,t} \qquad \forall t \qquad (6d)$$

$$\sum_s P_{i,j,s,t} h_s \le CF_{i,j} Cap_{i,j,t} \sum_s h_s \qquad \forall i,j,t \qquad (7a)$$

$$b_l \left| \sum_i S_{l,i} \theta_{i,s,t} \right| \le F_l^{\max} \qquad \forall l,s,t \qquad (8)$$

In GEP-11b, we observe that it will retire capacity when savings from salvage value and from avoiding FOM exceeds

- savings obtained via using the unit to displace energy & variable O&M from a more expensive unit over the entire planning horizon (always the case if the unit is never used), AND
- the capacity is not needed to satisfy reserves via (6d).

2.11 End effects

A key issue for multi-period planning is *end effects*. End effects refer to the difficulty of appropriately representing the influence of investment costs and operational costs at the end of the planning period, T_f. There are two problems to address:

1. *Remnant investment value*: The retirement year for many facilities occurs after the final year T_f. For these facilities, there is a value to the facility because it has remaining life, i.e., some of the investment paid has not yet been depreciated. If this value at $t=T_f$ is not subtracted from the objective, then the objective's investment portion is too high in the solution. This occurs whenever a unit is not retired on or before yr T_F.
2. *Remnant operational cost*: Because the simulation must be truncated at a particular final year T_f, operational costs after T_f are not included since those years are not simulated; then the objective's operational portion is too low in the solution.

Both above effects tend to bias decisions in favor of alternatives having low investment costs and high operational costs. Decisions in favor of low investment costs are more favored than they otherwise would be because

- ignoring the remnant investment value and the remnant operational cost
- means the objective function is overly-dominated by investment costs,
- and because it is minimizing costs, it chooses the investments with the lowest investment costs.

Decisions in favor of high operational costs are more favored than they otherwise would be because, since only a portion of the plant's operational life is considered, the high operational costs don't look so high.

We may address the problem of remnant investment value, for units having remaining life at the end of the planning horizon, by subtracting their $t=T_f$ salvage value from the objective. Assuming there are no further investments after T_f, the *Remnant Investment* at $t=T_f$ is the sum of the net present value of salvage values of units with remaining life at $t=T_f$, which will be:

$$\text{RemnantInvestment}=\zeta^{T_f-1}\left\{\sum_i\sum_j SV_{i,j,T_f}\,Cap_{i,j,T_f}\right\} \qquad (10)$$

The *RemnantInvestment* of (10) should be subtracted from the objective function (1c). Here, we assume the expression (1c) given above only sums across time $t=1,\ldots, t=T_f$. We make this explicit via (1d) below.

$$\begin{aligned}
\min \sum_{t=1}^{T_f}\zeta^{t-1}&\left\{\underbrace{\sum_i\sum_j I_{i,j,t}\,Cap^{add}_{i,j,t}-SV_{i,j,t}\,Cap^{ret}_{i,j,t}}_{\text{Investment Costs Less Salvage Values for Retired Unit}}+\underbrace{\sum_i\sum_j OM^{fixed}_{i,j}\,Cap_{i,j,t}}_{\text{Fixed O\&M Costs}}\right\}\\
+\sum_{t=1}^{T_f}\zeta^{t-1}&\left\{\underbrace{\sum_i\sum_j FC_{i,j,t}\,H_{i,j}\sum_s P_{i,j,s,t}\,h_s}_{\text{Fuel Costs}}+\underbrace{\sum_i\sum_j OM^{var}_{i,j}\sum_s P_{i,j,s,t}\,h_s}_{\text{Variable O\&M Costs}}\right\}\\
-\zeta^{T_f-1}&\left\{\sum_i\sum_j SV_{i,j,T_f}\,Cap_{i,j,T_f}\right\}
\end{aligned} \qquad (1d)$$

This addresses the issue of remnant investment value, but it does not address the issue of remnant operational value.

To address the issue of remnant operational value, we will <u>extend the simulation time beyond the planning horizon</u>. A multiperiod formulation with extended planning horizon is based on the observation that end effects increase their influence as the final year gets closer to the present year.

For example, if the planning horizon is 20 years, end effects have more influence in years 5-20 than they do in years 1-5. This observation leads to a very natural solution: extend the final year well-beyond the planning horizon. Therefore if the planning horizon is 20 years, we may run the optimization problem to 50 years; but the decisions for years 20-50 will be ignored. Alternatively, the additional years may be simulated but without load growth or further retirements, so that there is no need for additional investment.

A general guide for how far to extend the simulation beyond the planning horizon is that the final year of the simulation should exceed the final year of the planning horizon by the lifetime of the facility with the longest remaining life. This way, if an investment is made in the final year of the planning horizon, its full operational life can be evaluated; the fact that the extended time equals the lifetime of the facility with the longest life guarantees that the full life of all investments will be evaluated. We denote the extended last simulation year as $t=T_{MaxLife}$.

The production costs are given by FOM, VOM, and fuel costs. Then, they are expressed as

$$\text{RemnantOpCosts}=\sum_{t=T_f+1}^{T_f+T_{MaxLife}} \zeta^{t-1} PC(t) =$$

$$\sum_{t=T_f+1}^{T_f+T_{MaxLife}} \zeta^{t-1} \left\{ \underbrace{\sum_i \sum_j OM_{i,j}^{fixed} Cap_{i,j,t}}_{\text{Fixed O\&M Costs}} + \underbrace{\sum_i \sum_j FC_{i,j,t} H_{i,j} \sum_s P_{i,j,s,t} h_s}_{\text{Fuel Costs}} + \underbrace{\sum_i \sum_j OM_{i,j}^{\text{var}} \sum_s P_{i,j,s,t} h_s}_{\text{Variable O\&M Costs}} \right\} \tag{11}$$

Computing Remnant Operational Costs this way, however, requires computing over all of the times t, blocks s, technologies j, and buses i. This will increase computation time significantly. Instead, we make a rough assumption: each year’s operational costs remain the same as the last year of the planning horizon, i.e., $PC(t)=PC(T_F)$ for all $t=T_F+1, \ldots, T_{MaxLife}$. Then, equation (11) becomes:

$$\text{RemnantOpCosts}=\sum_{t=T_f+1}^{T_f+T_{MaxLife}} \zeta^{t-1} PC(T_F) = PC(T_F)\sum_{t=T_f+1}^{T_f+T_{MaxLife}} \zeta^{t-1} \tag{12}$$

We have already said that, with i as the discount rate, the discount factor is

$$\zeta^{t} = \frac{1}{(1+i)^{t}}$$

and so the term used in (12) is

$$\zeta^{t-1} = \frac{1}{(1+i)^{t-1}}$$

And the sum in (12) then just becomes a constant, i.e.,

$$\begin{aligned}\text{RemnantOpCosts}&=PC(T_F)\times \sum_{t=T_f+1}^{T_f+T_{MaxLife}} \zeta^{t-1}\\ &= PC(T_F)\times \sum_{t=T_f+1}^{T_f+T_{MaxLife}} \frac{1}{(1+i)^{t-1}} = PC(T_F)\times K\end{aligned} \tag{13}$$

Thus, the final objective function expression becomes (1e):

$$\begin{aligned}\min \sum_{t=1}^{T_f} \zeta^{t-1} &\left\{ \underbrace{\sum_i \sum_j I_{i,j,t} Cap_{i,j,t}^{add} - SV_{i,j,t} Cap_{i,j,t}^{ret}}_{\text{Investment Costs Less Salvage Values for Retired Unit}} + \underbrace{\sum_i \sum_j OM_{i,j}^{fixed} Cap_{i,j,t}}_{\text{Fixed O\&M Costs}} \right\}\\ + \sum_{t=1}^{T_f} \zeta^{t-1} &\left\{ \underbrace{\sum_i \sum_j FC_{i,j,t} H_{i,j} \sum_s P_{i,j,s,t} h_s}_{\text{Fuel Costs}} + \underbrace{\sum_i \sum_j OM_{i,j}^{\text{var}} \sum_s P_{i,j,s,t} h_s}_{\text{Variable O\&M Costs}} \right\}\\ - \zeta^{T_f-1} &\left\{ \sum_i \sum_j SV_{i,j,T_f} Cap_{i,j,T_f} \right\} + PC(T_F)\times \sum_{t=T_f+1}^{T_f+T_{MaxLife}} \frac{1}{(1+i)^{t-1}}\end{aligned} \tag{1e}$$

where

$$PC(T_F) = \underbrace{\sum_i \sum_j OM_{i,j}^{fixed} Cap_{i,j,T_F}}_{\text{Fixed O\&M Costs}} + \underbrace{\sum_i \sum_j FC_{i,j,T_F} H_{i,j} \sum_s P_{i,j,s,T_F} h_s}_{\text{Fuel Costs}} + \underbrace{\sum_i \sum_j OM_{i,j}^{\text{var}} \sum_s P_{i,j,s,T_F} h_s}_{\text{Variable O\&M Costs}}$$

There have been other ways suggested to handle end-effects, as indicated in [16, 17]. One way is called the primal equilibrium method and the other is called the dual equilibrium method. The basic idea of the primal equilibrium method is that

- The costs of the objective function are written as infinite series in time, from t=1,…,∞;
- A relation is imposed on all decision variables *x*, which requires that, beyond a certain time *T*, the value in one time period relates to the value in another time period, according to

  $$x(T+t) = \lambda \; x(T+t-1)$$

  where λ is the annual demand growth. Doing so means the infinite series beyond T can be approximated with a single term, so that the objective function can be modeled as before, for t=1,…,T and just a single term for the infinite horizon after that.

The basic idea of the dual equilibrium is similar to that of the primal equilibrium, except the equilibrium condition is imposed on the dual variables (prices) instead of the primal variables (capacity additions).

A comparative study was performed on a 5-node representation of the US generation portfolio to observe the influence of end effects using different methods, including (a) truncation to 40 years; (b) extended to 120 years; (c) include salvage values; (d) dual equilibrium [18]. Results are illustrated below.

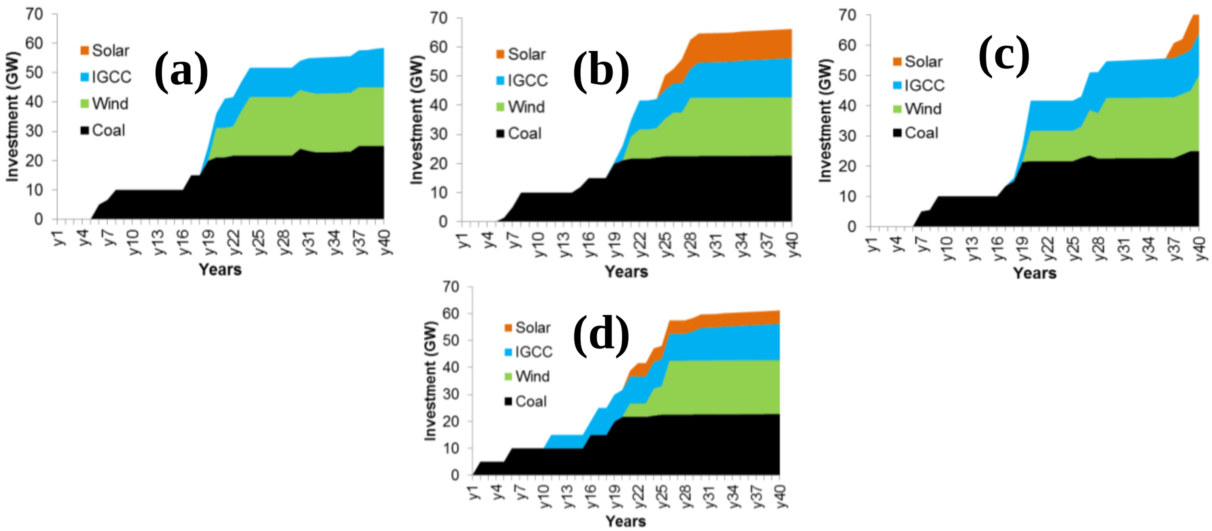


Figure 2-12: (a) Truncated; (b) Extended; (c) Salvage value; (d) Dual equilibrium

We take the extended 120 year simulation (Plot (b)) to be the basis of comparison. Comparing Plot (b) to Plot (a), it is clear that the truncated approach misses solar investment because, although solar operational costs are obviously low, their investment costs (at the time of this study), were extremely high. The other two approaches, which implement a method to account for end effects, tend to do better with the solar investments.

3 Overview of commercial tools

Some commercially available expansion planning tools include EGEAS, GEM, STRATEGIST, PLEXOS, and SERVM.

The New Zealand Electricity Commission developed GEM as a long range generation capacity planning model. It is formulated as a mixed integer programming problem (MIP). The computer code is written using the GAMS optimization software and the model is solved with CPLEX.

The EPRI product EGEAS (**E**lectric **G**eneration **E**xpansion **A**nalysis **S**ystem) is a modular production cost and generation expansion software package which employs dynamic programming algorithm to form candidate portfolios from identified alternatives meeting a capacity planning constraint. It also has modules which accommodate demand-side management options and facilitate development of environmental compliance plans. Some of the key functions of EGEAS are asset retirement evaluation, emission evaluation from new plants, and Scenario analysis for various generation options. EGEAS is widely used by many utilities and regulators.

The Ventyx product STRATEGIST is composed of multiple application modules incorporating all aspects of utility planning and operations. This includes forecasted load modeling, marketing and conservation programs, production cost calculations including the dispatch of energy resources, benefit-cost (B/C) ratios calculation for different alternatives, capital project modeling, financial and rate impacts evaluation, and analysis of long-range rate strategy and the implications of utility plans on customer classes.

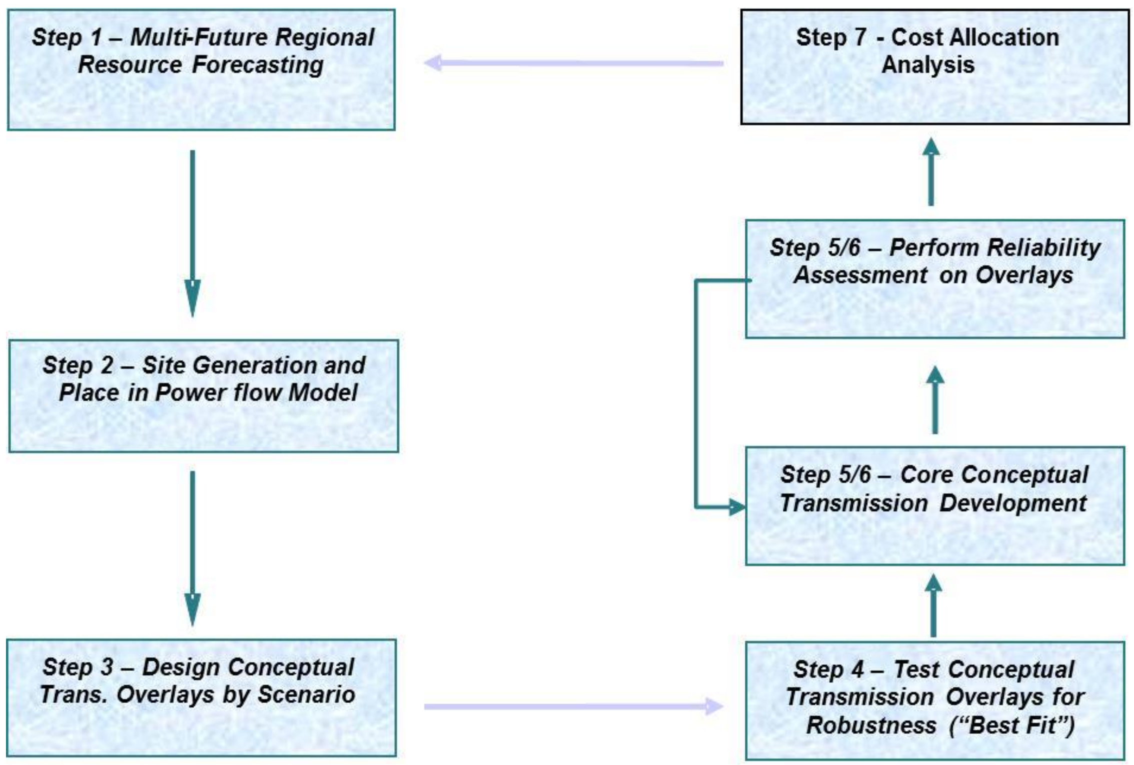
PLEXOS, from Energy Exemplar, is a versatile software system that finds optimal combinations of new generation units, unit retirements and transmission system upgrades on a least-cost basis over a long-term planning horizon. PLEXOS in itself does not incorporate the optimization engine but rather produces optimization code that can be read by an external solver such as CPLEX or MOSEK.

We provide an overview of EGEAS, adapted from [19].

The EGEAS computer model was developed by researchers at MIT under funding from the Electric Power Research Institute (EPRI). EGEAS can be run in both the expansion optimization and the production simulation modes. Uncertainty analysis, based on automatic sensitivity analysis and data collapsing via description of function estimation, is also available. A complete description of the model can be found in [20].

The production simulation option consists of production cost/reliability evaluation for a specified generating system configuration during one or more years. Probabilistic production cost/reliability simulation is performed using a load duration curve based model. Customer load and generating unit availability are modeled as random variables to reflect demand fluctuations and generation forced outages. Two algorithmic implementations are available: an analytic representation of the load duration curve (cumulants) and a piecewise linear numerical representation.

The below illustrates the MISO planning process. Step 1 is based on use of EGEAS.



Appendix – Load Duration Curves

A critical issue for planning is to identify the total load level for which to plan. One extremely useful tool for doing this is the so-called load duration curve, which is formed as follows. Consider that we have obtained, either through historical data or through forecasting, a plot of the load vs. time for a period T, as shown in Fig. A1 below.

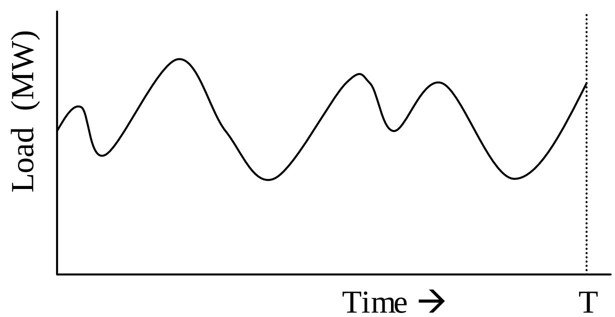


Fig. A1: Load curve (load vs. time)

Of course, the data characterizing Fig. A1 will be discrete, as illustrated in Fig. A2.

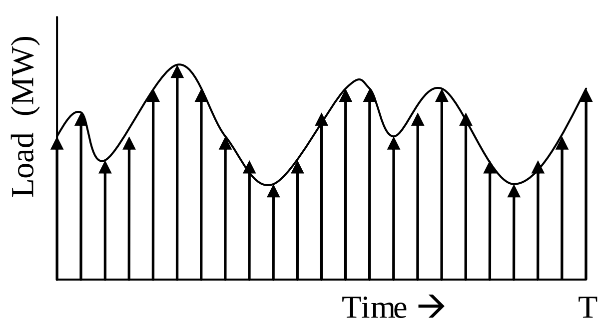


Fig. A2: Discretized Load Curve

We now divide the load range into intervals, as shown in Fig. A3.

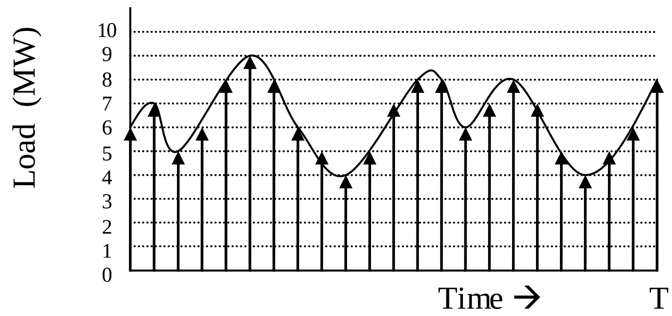


Fig. A3: Load range divided into intervals

This provides the ability to form a histogram by counting the number of time intervals contained in each load range. In this example, we assume that loads in Fig. A3 at the lower end of the range are "in" the range. The histogram for Fig. A3 is shown in Fig. A4.

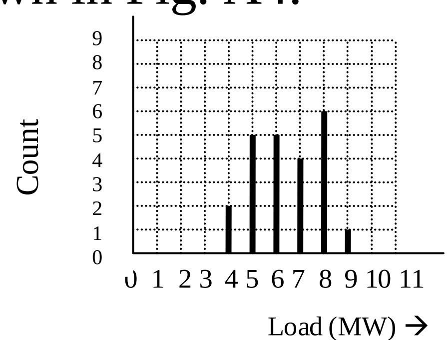


Fig. A4: Histogram

Figure A4 may be converted to a probability mass function, pmf (which is the discrete version of the probability density function, pdf) by dividing each count by the total number of time intervals, which is 23. The resulting plot is shown in Fig. A5.

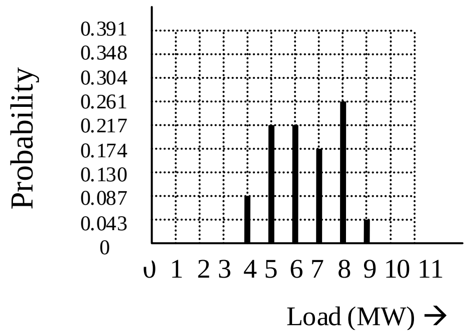


Fig. A5: Probability mass function

Like any pmf, the summation of all probability values should be 1, which we see by the following sum:

0.087+0.217+0.217+0.174+0.261+0.043=0.999

(It is not exactly 1.0 because there is some rounding error). The probability mass function provides us with the ability to compute the probability of the load being within a range according to:

$$\Pr(Load\ withinRange) = \sum_{L\ in\ Range} \Pr(Load = L) \qquad (2)$$

We may use the probability mass function to obtain the cumulative distribution function (CDF) as:

$$\Pr(Load \geq Value) = \sum_{L \geq Value} \Pr(Load = L) \qquad (3)$$

From Fig. A5, we obtain:

$$\Pr(Load \geq 1) = \sum_{L \geq 1} \Pr(Load = L) = 1.0$$

$$\Pr(Load \geq 2) = \sum_{L \geq 2} \Pr(Load = L) = 1.0$$

$$\Pr(Load \geq 3) = \sum_{L \geq 3} \Pr(Load = L) = 1.0$$

$$\Pr(Load \geq 4) = \sum_{L \geq 2} \Pr(Load = L) = 1.0$$

$$\Pr(Load \geq 5) = \sum_{L \geq 5} \Pr(Load = L)$$
$$= 0.217 + 0.217 + 0.174 + 0.261 + 0.043 = 0.912$$

$$\Pr(Load \geq 6) = \sum_{L \geq 6} \Pr(Load = L)$$
$$= 0.217 + 0.174 + 0.261 + 0.043 = 0.695$$

$$\Pr(Load \geq 7) = \sum_{L \geq 6} \Pr(Load = L) = 0.174 + 0.261 + 0.043 = 0.478$$
$$\Pr(Load \geq 8) = \sum_{L \geq 8} \Pr(Load = L) = 0.261 + 0.043 = 0.304$$
$$\Pr(Load \geq 9) = \sum_{L \geq 9} \Pr(Load = L) = 0.043$$
$$\Pr(Load \geq 10) = \sum_{L \geq 10} \Pr(Load = L) = 0$$

Plotting these values vs. load results in the CDF of Fig. A6.

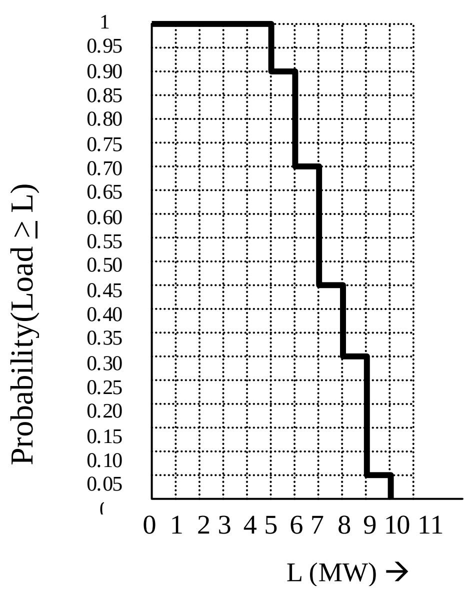


Fig. A6: Cumulative distribution function

The plot of Fig. A6 is often shown with the load on the vertical axis, as given in Fig. A7.

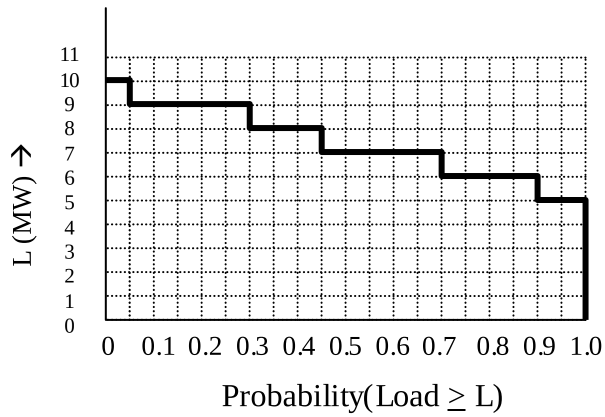


Fig. A7: CDF with axes switched

If the horizontal axis of Fig. A7 is scaled by the time duration of the interval over which the original load data was taken, T, we obtain the *load duration curve*. This curve provides the number of time intervals that the load equals, or exceeds, a given load level. For example, if the original load data had been taken over a year, then the load duration curve would show the number of hours out of that year for which the load could be expected to equal or exceed a given load level, as shown in Fig. A8.

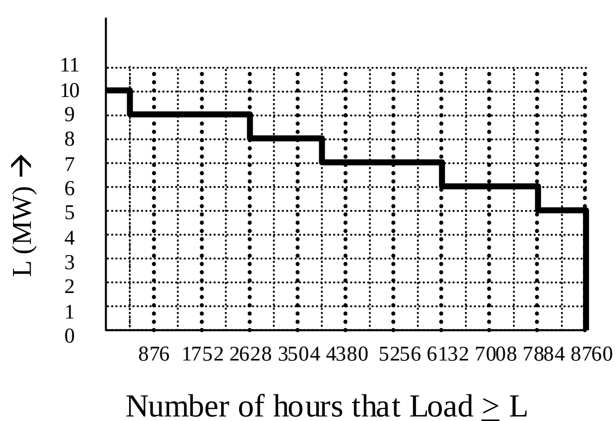


Fig. A8: Load duration curve

Load duration curves are useful in a number of ways.

- They provide guidance for judging different alternative plans. For example, one plan may be satisfactory for loading levels of 90% of peak and less. One sees from Fig. A8 that such a plan would be unsatisfactory for 438 hours per year (or 5% of the time).
- They identify the base load. This is the value that the load *always* exceeds. In Fig. A8, this value is 5 MW.
- They provide convenient calculation of energy, since energy is just the area under the load duration curve. For example, Fig. A9 shows the area corresponding to the base load energy consumption, which is 5MWx8760hr=43800 MW-hrs.

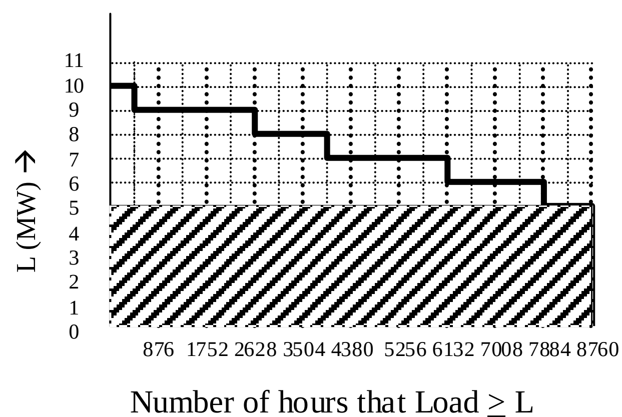


Fig. A9: Area corresponding to base load energy consumption

- They allow illustration of generation commitment policies and corresponding yearly unit energy production, as shown in Fig A10, where we see that the nuclear plant and coal plant #1 are base loaded plants, supplying 26280 MWhrs and 17520 MWhrs, respectively. Coal plant #2 and natural gas combined cycle (NGCC) plant #1 are the mid-range plants, and combustion turbine gas plant #1 is a peaker.

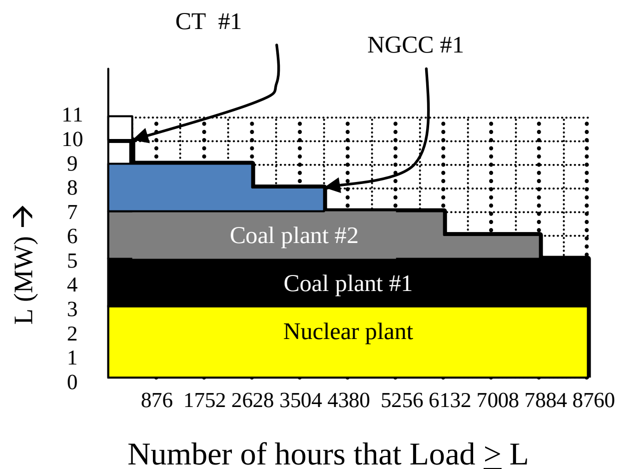


Fig. A10: Illustration of Unit commitment policy

Load duration curves are also used in reliability, production costing, and expansion planning programs in computing different reliability indices.

---

[1] North American Reliability Corporation (NERC), "2020 Long-term reliability assessment," Dec., 2020, [Online]. Available: https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_LTRA_2020.pdf.
[2] North American Reliability Corporation (NERC), "2020 Long-term reliability assessment," Dec., 2023, [Online]. Available: https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_LTRA_2023.pdf.
[3] See "Planning year 2013-2014: Wind Capacity Credit," https://www.midwestiso.org/Library/Repository/Study/LOLE/2013%20Wind%20Capacity%20Report.pdf

[4] MidContinent Independent System Operator (MISO), “Planning Year 2021-2022: Wind and Solar Capacity Credit,” January, 2021. [Online]. Available: https://cdn.misoenergy.org/2021%20Wind%20&%20Solar%20Capacity%20Credit%20Report503411.pdf/
[5] MidContinent Independent System Operator (MISO), “Planning Year 2023-2024: Wind and Solar Capacity Credit Report," Accessed 2/22/2024, Available: https://cdn.misoenergy.org/2023%20Wind%20and%20Solar%20Capacity%20Credit%20Report628118.pdf.
[6] A. Pereira and J. Saraiva, “Generation expansion planning in competitive electricity markets,” Proc. of the Power Tech Conference, 2007.
[7] S. Kannan, S. Baskar, J. McCalley, and P. Murugan, “Application of NSGA-II Algorithm to Generation Expansion Planning,” IEEE Transactions on Power Systems, Vol. 24, Issue 1, 2009, pp. 454-461.
[8] Alizadeh, B.; Jadid, S.: 'Reliability constrained coordination of generation and transmission expansion planning in power systems using mixed integer programming', IET Generation, Transmission & Distribution, 2011, 5, (9), p. 948-960, DOI: 10.1049/iet-gtd.2011.0122. IET Digital Library, https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2011.0122.
[9] J. A. Bloom, “Solving an Electricity Generating Capacity Expansion Planning Problem by Generalized Benders’ Decomposition,” Operations Research, vol. 31, no. 1, pp. 84–100, Jan-Feb. 1983.
[10] A. P. Sanghvi and I. H. Shavel, “Incorporating explicit loss-of-load probability constraints in mathematical programming models for power system capacity planning,” Int. J. Electrical Power &Energy Syst, vol 6, no. 4, pp. 239-247, Oct. 1984.
[11] J. Choi, T. Tran, A. A. El-Keib, R. Thomas, H. S. Oh, and R. Billinton, “A method for transmission system expansion planning considering probabilistic reliability criteria,” IEEE Transactions on Power Systems, vol. 20, no. 3, pp. 1606–1615, Aug. 2005, doi: 10.1109/TPWRS.2005.852142.
[12] J. H. Roh, M. Shahidehpour, and L. Wu, “Market-based generation and transmission planning with uncertainties,” IEEE Transactions on Power Systems, vol. 24, no. 3, pp. 1587–1598, 2009, doi: 10.1109/TPWRS.2009.2022982.
[13] S. Dehghan, N. Amjady, and A. J. Conejo, “Reliability-Constrained Robust Power System Expansion Planning,” IEEE Transactions on Power Systems, vol. 31, no. 3, pp. 2383–2392, May 2016, doi: 10.1109/TPWRS.2015.2464274.
[14] B. Frew, G. Stephen, D. Sigler, J. Lau, W. B. Jones, and A. Bloom, “Evaluating Resource Adequacy Impacts on Energy Market Prices Across Wind and Solar Penetration Levels,” The Electricity Journal, vol. 32, no 8, pp. 2019.
[15] L. C. Da Costa, F. S. Thome, J. D. Garcia, and M. V. F. Pereira, “Reliability-Constrained Power System Expansion Planning: A Stochastic Risk-Averse Optimization Approach,” IEEE Transactions on Power Systems, vol. 36, no. 1, pp. 97–106, Jan. 2021, doi: 10.1109/TPWRS.2020.3007974.
[16] R. C. Grinold, “Model Building Techniques for the Correction of End Effects in Multistage Convex Programs,” *INFORMS*, Vol. 31, No. 3, pp. 407-431, May-Jun., 1983
[17] S. C. Walkar, “Evaluating End Effects for Linear and Integer Programs using Infinite-Horizon Linear Programming,” *PhD Dissertation, Naval Postgraduate School*, March 1995
[18] V. Krishnan and J. McCalley, “Building foresight in long-term infrastructure planning using end-effect mitigation models,” DOI 10.1109/JSYST.2015.2410758, *IEEE Systems Journal*, Vol. 11, Issue 4, 2017, pp. 2040-2051.
[19] International Atomic Energy Agency, “EXPANSION PLANNING FOR ELECTRICAL GENERATING SYSTEMS: A Guidebook,” 1984.
[20] ELECTRIC POWER RESEARCH INSTITUTE, Electric Generation Expansion Analysis Systems — Vol. 1: Solution Techniques, Computing Methods, and Results, Rep. EPRI EL-256K1982).