Augmented screening curve analysis of thermal generation capacity additions with increased renewables, ancillary services, and carbon prices

Ross Baldick, Heejung Park, Duehee Lee
Department of Electrical and Computer Engineering
The University of Texas at Austin
August 2011
Abstract

A “traditional” screening curve approach can be useful to estimate optimal additions to dispatchable generation capacity for a target year. However, this approach does not consider the capacity needed for reserves and other online, but unloaded, capacity. With increasing penetration of renewables, at least two issues need to be considered. First, the load-duration curve used in a traditional screening approach should be replaced with a net load-duration curve, where net load is defined to be the load minus intermittent renewable production. In principle, this simply requires evaluation of the cumulative distribution function of the net load for the relevant future demand and wind scenarios. Second, with significantly increased intermittent renewables, there is an increase in the amount of capacity needed to provide ancillary services, and both the capital and operating costs of this capacity should be considered in order to fully evaluate the costs of meeting the net load plus ancillary services. The presentation will approach these issues, and include consideration of the effect of carbon prices.

Work in progress! Critical feedback appreciated!
Outline

(i) Generation capital planning,
(ii) Review of basic screening curve approach,
(iii) Extensions,
(iv) Ancillary services,
(v) Further extensions and limitations,
(vi) Conclusion.
1 Generation capital planning

- Range of approaches to generation capital planning:
  - from “screening curve,”
  - to mixed integer programming.
- Different approaches have different roles and levels of detail:
  - cost-based versus market-based,
  - representation of transmission,
  - representation of detailed operations,
  - representation of lumpiness of capital additions.
Generation capital planning, continued

- Goal here is to understand implications for cost-based thermal generation expansion due to:
  - integration of intermittent renewables, and/or
  - imposition of carbon price.
- Use stylized model that ignores:
  - transmission (or assume proxy included in generation costs),
  - many details of operation,
  - lumpiness of capital and operating decisions.
- Aim is to represent:
  - randomness of intermittent renewables,
  - case where capital planning is partly driven by need to provide reserves and other ancillary services (AS),
  - “economically adapted” thermal generation portfolio, so as to anticipate likely changes to current thermal portfolio.
2 Review of basic screening curve approach

2.1 Background

- Early method for understanding capital additions to minimize total operating and capital costs [1]:
  - focuses on future target year,
  - assumes annualized representation of capital costs,
  - ignores integrality of capital and operating decisions, including minimum generation limits.

- Standard formulation:
  - ignores “no-load operating costs,”
  - ignores “start-up costs,” and other inter-temporal issues,
  - finds economically adapted capacity based on costs, does not represent existing capacity,
  - does not represent transmission unless included as proxy in generation costs,
  - does not represent maintenance outages.

- Extensions to issues such as storage are possible.
Background, continued

- What is value of screening curve, given the many approximations?
  - Most formal generation expansion models do not represent transmission,
  - Economically adapted generation levels for target future year can provide basis for detailed planning of expansion over horizon,
  - Can reformulate as an optimization problem to include more issues [2],
  - Cost-based results are guide to competitive market outcomes,
  - Can obtain analytic insights about the effects of varying wind penetration.
  - Can estimate cost of wind integration.
  - Can estimate long-term effects of carbon tax on thermal portfolio.

- Screening curve:
  2.2 Assumptions,
  2.3 Annual general costs,
  2.4 Load-duration curve,
  2.5 Optimal mix of capacity.
2.2 Assumptions

Generation:

- dispatchable,
- each technology is characterized by a simple model of capital and operating costs:
  - capital recovery cost \( c \) in \$/MW.year that represents the annual requirements for depreciation, taxes, and return on equity; and
  - a single representative variable operating cost \( v \) in \$/MWh.
- lumpiness of capital additions and of commitment decisions can be ignored,
- effects of forced outages of generators can be represented by appropriate derating factors on capacities,
- inter-temporal issues such as start-up costs are ignored.

Transmission:

- costs are negligible or included in generation costs, and
- no transmission congestion.

Load-duration curve: (the cumulative distribution function of the load) is known.
2.3 Annual generation costs

- Suppose that an amount of capacity of a particular technology was operated for a fraction $t_0$ of the year.
- The parameter $t_0$ is the annual capacity factor.
- Total cost per unit capacity per year for both capital and variable costs is $c + 8760 \times vt_0$ (units are in $$/MW\cdot year)$.
- Total cost per unit capacity per year is an increasing, affine function of the annual capacity factor $t_0$.

Fig. 1. Total cost per unit capacity versus capacity factor for three technologies.
2.3.1 Comparing technologies

• Basic insight in screening curve approach:
  – “peaking” generation (high variable cost, low capital cost) is cheapest when capacity used at low capacity factor,
  – “cycling” generation (medium variable cost, medium capital cost) is cheapest when capacity used at medium capacity factor, while
  – “baseload” generation (low variable cost, high capital cost) is cheapest when capacity used at high capacity factor.

• For particular case of Figure 1 with three generation technologies:
  – crossover between “peaking” and “cycling” occurs at threshold capacity factor 0.3, and
  – crossover between “cycling” and “baseload” occurs at threshold capacity factor 0.7.

• In general, crossover depends on relationship between $c$ and $v$ for each pair of technologies.
Comparing technologies, continued

• With multiple technologies there will be capacity factor thresholds between regions of optimality for each technology:
  – super-peaking internal combustion or other reciprocating engine,
  – peaking gas turbine (GT),
  – cycling combined-cycle gas turbine (CCGT),
  – baseload black coal,
  – baseload lignite, and
  – baseload nuclear.
Comparing technologies, continued

• For a given set of base fuel prices, imposition of carbon tax or cap and trade shifts the relative technology thresholds due to effective increase in variable cost of using fossil fuels:
  – GT/CCGT threshold would decrease because GT variable cost would increase more than CCGT variable cost,
  – CCGT/coal threshold would increase because CCGT variable cost would increase less than coal variable cost, and
  – coal/nuclear threshold would decrease because coal variable cost would increase.
  – less GT, more CCGT, less coal, and more nuclear in economically adapted portfolio, all other things equal.

• As we will see, the effect of onshore wind on thresholds is to increase “peakiness” of the “net load” (load minus wind) that is met by thermal system:
  – increase peaking in economically adapted portfolio, and
  – decrease baseload.
2.3.2 Comparing baseload and other technologies

- Assumptions:
  - 2.5% annual depreciation,
  - 35% corporate taxes, and
  - approximately 8% return on equity.
- Implies capacity recovery cost $c$ is approximately 0.15 of capital cost.
- Consider generic capital costs and heat rates based on US Energy Information Administration data [3].

<table>
<thead>
<tr>
<th>Technology</th>
<th>Capital cost (US$/kW)</th>
<th>Heat rate (Btu/kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear</td>
<td>3500</td>
<td>10,500</td>
</tr>
<tr>
<td>Coal</td>
<td>2000</td>
<td>9,000</td>
</tr>
<tr>
<td>Combined cycle</td>
<td>1000</td>
<td>7,000</td>
</tr>
<tr>
<td>Gas turbine</td>
<td>650</td>
<td>10,500</td>
</tr>
</tbody>
</table>

- Assume coal price is US$2.3/million Btu and that nuclear fuel is essentially free.
- About 1055 Joules per Btu, so coal price is around US$2.20/GJ.
Comparing baseload and other technologies, continued

- Even assuming 100% capacity factor:
  - combined cycle is lower cost overall than coal for gas prices below US$5.40 per million Btu, and
  - combined cycle is lower cost overall than nuclear for gas prices below US$6.10 per million Btu.

- Shale gas is likely to keep gas prices not significantly above approximately US$5 per million Btu.

- For reasons that will become clear in the rest of presentation, growth in onshore wind in ERCOT will result in typical capacity factor of new thermal capacity being well below 100%:
  - at lower capacity factors, combined cycle will always be cheaper than coal or nuclear overall at even higher gas prices,
  - so economically adapted new capacity would not be baseload.
2.3.3 Effect of carbon tax

- Carbon emissions are on the order of one tonne per MWh of coal generation and one-half tonne for gas:
  - a carbon tax on order of US$30 per tonne would make marginal cost of coal around the same as that of combined cycle, and
  - economically adapted new capacity would not be coal.

2.3.4 Implications

- Given combined effects of low natural gas prices, increase in onshore wind, and implementation of carbon tax or cap and trade, we will mostly focus on choice between two technologies:
  - peaking, and
  - cycling.
2.4 Load-duration curve

- Equivalent to a cumulative distribution function (CDF) of load considered to be a random variable:
  - annual load-duration curve considers distribution over a year,
  - ignores inter-temporal characteristics of demand,
  - vertical axis shows load, and
  - horizontal axis shows capacity factor $t_0 = (1 - \text{CDF})$.

Fig. 2. Example load-duration curve.
2.5 Optimal mix of capacity

- Imagine “peaking” and “cycling” generators each 1 MW capacity:
  - how many peaking generators should we use?
  - how many cycling generators should we use?
- Divide the 1000 MW peak load into 1 MW slices and consider the capacity factor $t_0$ of each slice.
- Choose technology for a slice based on the capacity factor.

![Fig. 3. Example load-duration curve.](image-url)
Optimal mix of capacity, continued

- Given the threshold of 0.3:
  - peaking is optimal for capacity factor less than 0.3 (corresponding to 360 “peaking” generators each of 1 MW capacity in Figure 4), and
  - cycling is optimal for capacity factor greater than 0.3 (corresponding to 640 “cycling” generators each of 1 MW capacity in Figure 4).

Fig. 4. Optimal mix of capacity.
Optimal mix of capacity, continued

- In practice, generators come in larger lumps:
  - six 60 MW capacity peaking generators, and
  - two 320 MW capacity cycling generators.
- More detailed planning model could decide optimal capacities and construction timetable given lumpiness.
3 Extensions

- Several extensions have appeared in the literature or are straightforward applications of the basic screening curve approach:
  3.1 representing unserved demand,
  3.2 including non-dispatchable generation,
  3.3 representing existing generation.
- Use ERCOT data to illustrate some of these extensions.
3.1 Unserved demand

- The load-duration concept comes from “traditional” utility requirement of meeting a fixed forecast essentially independent of cost (or of market price).
- “Value of lost load” (VOLL) for unserved demand can be introduced as a pseudo-generator with zero fixed costs and variable cost equal to VOLL [4]:
  - at sufficiently low capacity factor, prefer to not serve demand rather than install peaker.
- Interruptible load having a reservation price and an exercise price could also be represented:
  - similar to generator cost curve.
- (Not clear how to represent varying value of demand over time horizon.)
3.2 Non-dispatchable generation

- Typical wind and solar photovoltaic renewable resources *cannot* be incorporated directly as a generation resource in the screening curve approach, since they are not dispatchable and therefore cannot be called upon to meet given demand.
- Instead, adjust load-duration curve to represent non-dispatchable resources:
  - assuming an exogenous specification of wind capacity.
- Define “net load” to be load minus non-dispatchable resources.
- Naive net load-duration curve for future scenario:
  - utilize several years of time series data for load and non-dispatchable resource,
  - scale load and non-dispatchable resource time series for future target year scenario,
  - subtract future non-dispatchable from future load data to obtain net load,
  - estimate cumulative distribution function (CDF) to obtain net load-duration curve.
3.2.1 Scaling of non-dispatchable generation

- Naive scaling approach may not appropriately represent averaging out of variability over short time scales:
  - particularly relevant for intermittent renewables such as wind and solar photovoltaic,
  - two windfarms have relatively less short-term variability than one,
  - may need filtering of time series data rather than simple scaling.

- We will use naive approach here!
  - Considering appropriate way to “filter” time series in other research.

- Focus on wind because of relevance to ERCOT.
Scaling of non-dispatchable generation, continued

Fig. 6. Power spectral density of ERCOT wind farms.
3.2.2 Net load-duration curve

- For given net load-duration curve, screening curve approach will provide economically adapted mix of capacity for future target year, given assumed penetration of wind.
- In systems where wind blows mostly off-peak, net load-duration will be “peakier” than load-duration.

Load and net load (MW)

\[ t_0 = (1 - CDF) \]

Fig. 7. Load-duration (thin line) and net load-duration (thick line) curve.
3.2.3 Optimal mix with non-dispatchable generation

- Addition of non-dispatchable generation will change optimal mix of capacity:
  - on-shore wind in ERCOT blows more off-peak than on-peak, so optimal mix shifts away from baseload and cycling and towards peaking generation.
  - solar photovoltaic produces closer to time of peak, so will shift mix away from cycling.
Optimal mix with non-dispatchable generation, continued

- Consider addition of on-shore wind resources.
- Given the threshold of 0.3 and example net load-duration curve:
  - peaking is optimal for 450 MW (compared to 360 MW previously), and
  - cycling is optimal for 500 MW (compared to 640 MW previously).

Fig. 8. Optimal mix based on net load-duration curve.
3.3 Existing generation

• Typical system has an existing portfolio of generation.
• Two possibilities for results of screening curve:
  (i) optimal capacity of each technology exceeds existing capacity, or
  (ii) optimal capacity of some technology (for example, baseload) is
   less than existing capacity.

3.3.1 Optimal capacity more than existing

• Difference between optimal and existing is indicative of needed
  incremental construction over planning horizon.
3.3.2 Optimal capacity less than existing

- Suppose optimal *baseload* from screening curve is less than existing baseload capacity and that:
  - retirements of baseload capacity are exogenous, and
  - operating cost of existing baseload is no more than operating cost of new non-baseload technologies.

- Then subtract derated capacity of baseload generation from load-duration curve:
  - yields residual load-duration curve,
  - if negative values, set residual load value to zero,
  - planned maintenance would presumably overlap with some of the times when baseload capacity exceeded load.

- Apply screening approach to residual load-duration curve:
  - only considering non-baseload generator technologies, under assumption that no baseload should be added.
Optimal capacity less than existing, continued

- Baseload coal is already occasionally operating at minimum during off-peak in markets such as Electric Reliability Council of Texas (ERCOT) and Australian NEM:
  - Increases in renewables will reduce need for new baseload, although some older baseload may be repowered or replaced.
  - Primary new technology addition decision is between “cycling” and “peaking.”
Optimal capacity less than existing, continued

Fig. 9. Net load-duration curve for ERCOT in 2009-2010 with baseload capacity shown.
Optimal capacity less than existing, continued

- Case that existing baseload operating cost exceeds operating cost of new generation is more difficult to handle:
  - could occur with large enough carbon tax,
  - need to model economic dispatch and endogenous retirement decisions.

- Case that optimal peaking capacity is less than existing peaking is more difficult to handle:
  - addition of intermittent renewables more likely to increase need for peaking capacity, decrease need for baseload capacity, so
  - case of existing peaking capacity exceeding optimal peaking may not be significant in practice.
3.4 ERCOT example

- ERCOT load growth is around 2% per year:
  - consider scale-up of 2009–2010 load by 20% to represent circa 2020.
- ERCOT has transmission expansion planned to accommodate a total of around 18 GW of wind in West Texas:
  - wind capacity around 10 GW currently,
  - consider scale-up of 2009–2010 wind to around 20 GW of capacity.
ERCOT example, continued

Fig. 10. Load-duration and net load-duration curves for ERCOT, with 2009–2010 load scaled-up by 20%.
ERCOT example, continued

Fig. 11. Load-duration and net load-duration curves for ERCOT. Existing baseload thermal capacities shown.
**ERCOT example, continued**

- As discussed previously, likely to be no significant need for new baseload generation, except to replace or repower retired capacity.
- For US$5 per million Btu gas and previous costs, threshold between combined cycle and gas turbine is approximately 0.34, corresponding to $8760 \times 0.34 = 3000$ hours.
- With 20GW of wind, need approximately:
  - 21.7 GW of combined cycle, compared to about 22 GW currently,
  - 36.5 GW of gas turbine, compared to about 15 GW of steam turbine and 3 GW of oil and gas turbine currently.
- At higher gas costs, would need relatively more combined cycle and relatively less peaker.
- Possible alternatives to peakers include storage and demand response:
  - main conclusion is that load growth and wind growth in ERCOT result in increase in need to cope with peaks, not in need for increased baseload.
4 Ancillary services

- The screening curve approach does not directly represent the “excess” online capacity needed for Ancillary Services (AS) and other issues.
- There is more to meeting load (or net load) than just load-duration issues:
  - net load level of 950 MW having 200 MW of forecast intermittent renewable generation might require 200 MW of reserves, while
  - net load level of 950 MW having 0 MW of forecast intermittent renewable generation might require only 100 MW of reserves.
- Representation of capacity needed for AS is more critical in the presence of intermittent renewables because:
  - increased intermittent renewables will necessitate more reserves, and
  - increased intermittent renewables will involve less online dispatchable generation, which will be called upon to provide a greater fraction of its online capacity for AS.
- Need for “excess” online capacity for ancillary services will further shift the optimal mix of capacity away from baseload and cycling and towards peaking generation.
- How to estimate the optimal capacity mix in this case?
4.1 Augmented net load-duration curve

- Instead of considering net load, consider “augmented net load,” the sum of:
  - net load, plus
  - “excess” online capacity needed for AS.

Augmented net load and net load (MW)

![Diagram showing Augmented net load and net load (MW)]

Fig. 12. Augmented net load-duration (thick line) and net load-duration (thin line) curve.
4.2 Additional assumptions

- For generation:
  - at any given time, each online generator contributes the same fraction of its online capacity to:
    - producing energy, and
    - providing “excess” online capacity for AS such as (up) regulation and spinning reserves;
  - rest of capacity is available for non-spinning reserves,
  - each technology is characterized by a simple model of cost:
    - capital recovery cost $c$ in $$/MW.year,
    - no-load operating cost per unit capacity $\ell$ in $$/h per MW of capacity, and
    - a single representative variable operating cost $v$ in $$/MWh for production above zero.

- Can again estimate total costs versus capacity factor, assuming known probabilistic relationships between load, reserves, and intermittent renewables.
Additional assumptions, continued.

- For each particular level $t \in [0, 1]$ of the augmented net load–duration curve, the probability density function of the distribution of the net load as a fraction $e \in [0, 1]$ of the augmented net load is assumed to be $f(t, e)$:
  - for each $t \in [0, 1]$ and each $e \in [0, 1]$, a fraction $e$ of online capacity is needed for energy,
  - assumption on generators is that each generator uses exactly this fraction $e$ of its online capacity to produce energy,
  - ignore contribution to operating costs due to actually deploying this excess capacity.

- Evaluation of probability density function $f$ requires analysis of the need for AS:
  - frequency regulation and reserves requirements due to load and thermal generation based on, for example, short-term variability/uncertainty of load and largest online generation, and
  - additional reserves due to intermittent renewables based on, for example, statistical information about conditional probability of wind die-off.
4.3 Cost of operation

- Estimate the expected cost per unit capacity of operating each technology to provide both energy and “excess” online capacity at capacity factor $t_0$:
  - if the fraction of capacity for energy is $e$ then operating cost per unit capacity per hour is $(\ell + ve)$,
  - at each particular capacity level $t \in [0, t_0]$, take the expected value of $(\ell + ve)$:
    \[
    \int_{e=0}^{1} (\ell + ve) f(t, e) de = \ell + v \int_{e=0}^{1} ef(t, e) de.
    \]
  - The expected operating cost per unit capacity per year of providing the appropriate fractions at the capacity factor $t_0$ is obtained by integrating this expression from zero to $t_0$ and multiplying by 8760 hours:
    \[
    8760 \times \int_{t=0}^{t_0} (\ell + v \int_{e=0}^{1} ef(t, e) de) dt = 8760[\ell t_0 + v \int_{t=0}^{t_0} \int_{e=0}^{1} ef(t, e) de dt].
    \]
- These operating costs involve the no-load operating cost incurred for a total of $8760 \times t_0$ hours together with the variable operating costs during these hours corresponding to the fraction of energy production.
Cost of operation, continued.

- The expected total cost per unit capacity per year are:

\[
c + 8760[\ell t_0 + v \int_{t=0}^{t_0} \int_{e=0}^{1} e f(t, e) de dt].
\]

- Unlike the simpler case considered for a traditional screening curve analysis, this expression is not linear in \(t_0\) unless, for example, \(f\) is independent of \(t\).
- Total cost per unit capacity per year is an increasing function of \(t_0\).
Cost of operation, continued

- Qualitatively, as the fraction of capacity used for reserves increases, the slope of the total cost curve will decrease.
- Tends to increase the threshold capacity factor between technologies.
- Further shifts the optimal mix of capacity away from cycling and towards peaking generation.

Total cost ($/MW.year)

![Graph showing Total cost per unit capacity versus capacity factor for two technologies with (thick lines) and without (thin lines) significant reserves.]

Fig. 13. Total cost per unit capacity versus capacity factor for two technologies with (thick lines) and without (thin lines) significant reserves.
4.4 Augmented screening analysis

- If significant “excess” online capacity is required, resources that have low capital costs will be even more attractive than baseload resources compared to considering net load–duration issues alone.
- This is because the model of provision of reserves considers the fixed capital costs and the no-load costs necessary to provide for “excess” online capacity:
  - with larger needs for reserves, variable cost of energy production becomes less important.
- As in the basic screening curve approach, for each capacity factor $t_0$, the lowest cost technology would be chosen to serve the corresponding capacity segment of the augmented net load-duration curve:
  - assuming each pair of total cost curves intersect at no more than one value of capacity factor.
Cost of operation, continued

- For these particular costs and (high) level of reserves, always use peaking for new generation.
- Less extreme change expected in practice.

Total cost ($/MW.year)

![Graph showing cost versus capacity factor for two technologies](image-url)

**Fig. 14.** Total cost per unit capacity versus capacity factor for two technologies considering reserves.
4.5 Optimal capacity

- Assuming previous costs, all additional 1050 MW of capacity would be peaking.

Augmented net load (MW)

0 1

Capacity factor \( t_0 = (1 - \text{CDF}) \)

Fig. 15. Augmented net load-duration curve.
4.6 ERCOT example

- Again consider scale-up of 2009–2010 load by 20% and 20 GW of wind.

Fig. 16. Residual augmented net load-duration and residual net load-duration curves for ERCOT.
• For simplicity, only consider spinning reserves:
  – in practice, non-spinning reserves are also required,
  – about 1 GW of spinning reserves provided by demand response in ERCOT.
• Again likely to be no significant need for baseload generation, except to replace retired capacity.
• For US$5 per million Btu gas and previous costs, but ignoring no-load operating costs, threshold between combined cycle and gas turbine remains 0.34, again corresponding to 3000 hours.
• With 20GW of wind, and assuming 1 GW of spinning reserves provided by demand response, would need approximately:
  – 23 GW of combined cycle, compared to about 22 GW currently,
  – 36.5 GW of gas turbine, compared to about 15 GW of steam turbine and 3 GW of oil and gas turbine currently.
• With non-zero no-load costs, threshold would increase, so more gas turbine and less combined cycle.
5 Further extensions and limitations

- The stylized model ignores many issues:
  - starting point for more detailed operations modeling including transmission,
  - model requires estimate of the dependence of the need for reserves on level of intermittent renewables and this might not be independent of inter-temporal issues such as ramping requirements,
  - starting point for detailed expansion plan between now and target year using dynamic programming.

- Transmission expansion would be difficult to include in detail because of:
  - lumpiness, and
  - complications of meshed system expansion.
Extensions and limitations, continued.

- It may be difficult to correctly represent the variation in provision of reserves due to economic dispatch of each technology:
  - significant issue if baseload and peaking were being compared,
  - less significant in the comparison between CCGT and GT, particularly if there are requirements to share reserves evenly across units,
  - correct representations would imply less cycling and more peaking capacity, since cycling capacity would not be used as much for reserves and peaking would be used more for reserves.
Extensions and limitations, continued.

- Some further details of operations may fit into the screening curve framework:
  - different types of reserves, given a known joint probability distribution of the needs for these different reserves:
    - non-spinning reserves could be incorporated as another type of needed capacity, with zero operating costs,
    - maintenance outages could also be approximately represented,
  - proxies to cost of start up costs, assuming that particular technologies are started weekly or daily,
  - demand shifting and storage.
6 Conclusion and ongoing work

- Basic screening curve approach to generation planning,
- Generalization to include issues with intermittent renewables:
  - net load-duration,
  - online capacity needed for reserves and other AS.
- Possible further extensions,
- Limitations.
- Ongoing work includes:
  - developing “scaling” approaches to better predict load-duration characteristics of future wind,
  - obtaining generic no-load cost data,
  - developing approaches to better predict AS requirements of future wind, and
  - considering non-spinning reserves.


Energy Information Administration. Cost and performance characteristics of new central station electricity generating technologies. Available from