

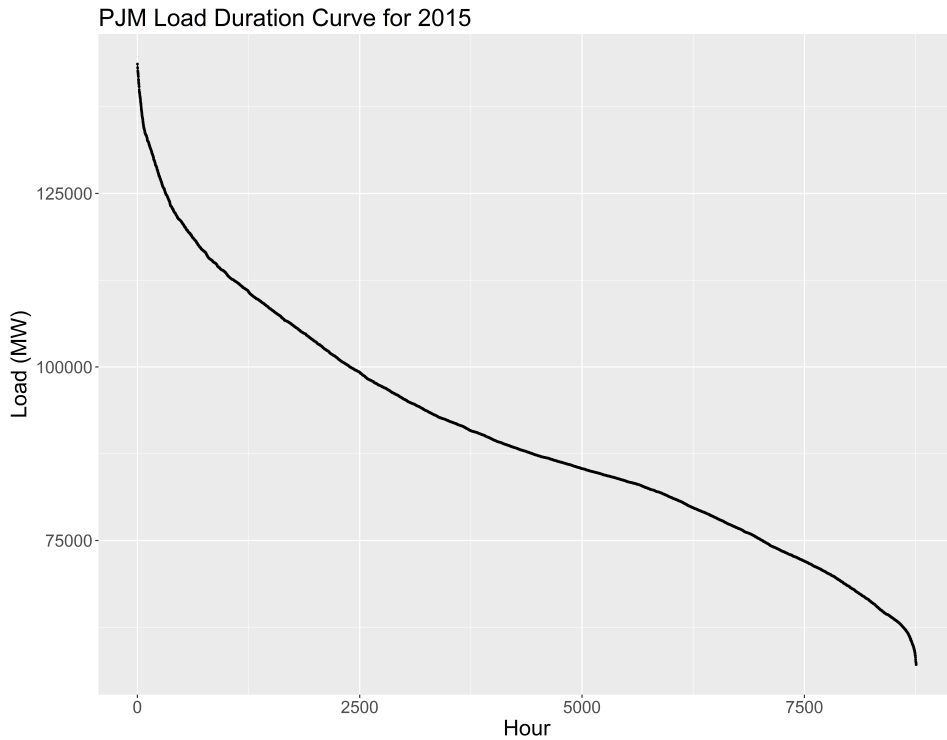
1. INTRODUCTION

The generation expansion planning (GEP) problem is used for long-term planning in the electricity industry in order to identify the optimal portfolio mix for some future period that minimizes the objective function, which are usually investment and operation cost, while meeting all of the imposed constraints (i.e., reliability, budget, etc.). A number of different model formulations are used to study this problem, including stochastic models and methods, which treat some of the parameters as uncertain. We propose to use a two-stage framework whereby the forced outage rate (FOR), which is the percentage of time a unit spends in unplanned outage, as a random variable (r.v.).

In the first of the two stages, the investment decisions are made (i.e., type, location, quantity). Then in the second stage, the random variables are realized, and operation decisions are made (i.e., how much to run each unit, etc.). A number of variables have been considered as stochastic in previous work, including: investment cost, demand, . . . , etc. However, to date no one has considered the reliability of the individual generator as a r.v. We propose to accomplish this in our work. However, as a first step we first discuss methods used to assess the reliability of a proposed power generation mix.

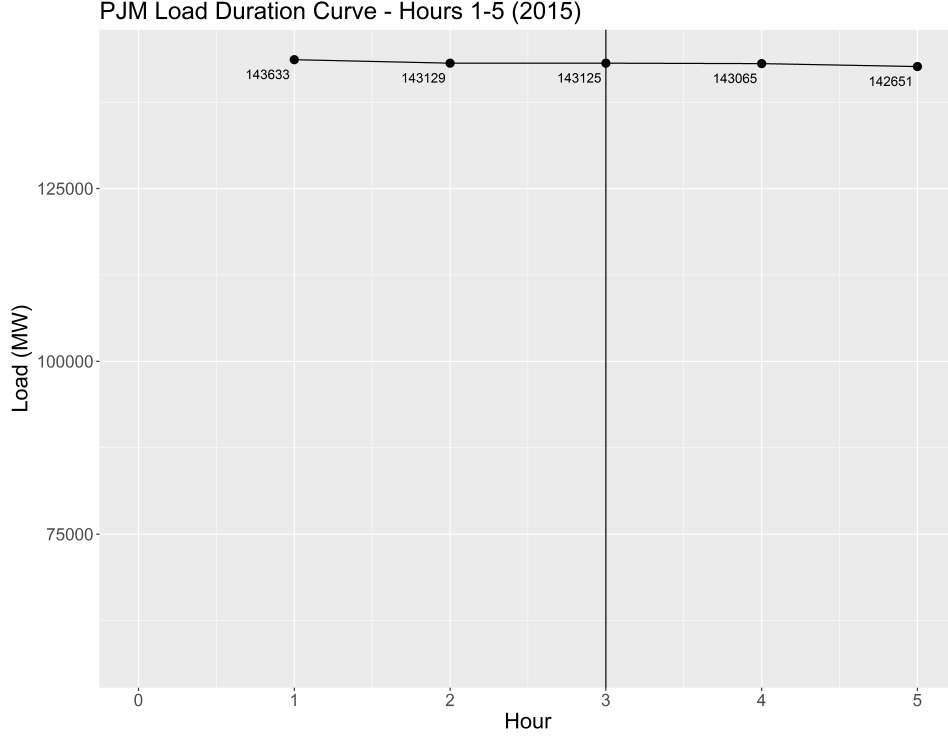
In Figure 1 we show the 2015 load duration curve (LDC) for the PJM regional transmission operator (RTO). The LDC is constructed by ordering the hourly load data from largest to smallest. The LDC can be used to determine the reliability metric, associated with a reliability percentage.

FIGURE 1. Load duration curve for PJM in 2015



If we zoom in on the highest load levels of the LDC (See Figure 2), we can see that if for the far left regio For example, if we had a reliability standard whereby the load can exceed available capacity three hours annually then if every year was like the data that we have for 2015, this would be equivalent to an installed capacity of 143,125 MW.

FIGURE 2. Load duration curve for PJM in 2015 - Maximum load



We can use the percentile (i.e., $1 - \frac{3}{8760}$) and the load value (i.e., 143,125) to determine the Poisson distribution uniquely determined by those values using the following code:

```
f2=function(lambda) ppois(x.value, lambda) - x.percentile
uniroot(f2, c(0, ub))
```

The function minimizes the error (i.e., distance between the expected probability of the reference value and the value from the LDC) over the search range looking for the λ value that minimizes the error between the expected probability and the actual probability.

- Most (if not all) power systems have a system level reliability standard (i.e., the percentage of time over a ten year period where demand exceeds available capacity)
- Each system has a load duration curve, an ordered list of electricity demand, from which using the exogeneously set reliability standard we can map the probability to a MW value

- System level reliability is an emergent property of generator reliability (might be an abuse of terms)
- Power plants have three states: on, off, or derate; though most reliability models simplify to two states (on and off)
- The probability of each state is determined by the forced outage rate (i.e., time spent in unplanned outage)
- Planners are interested in maintaining the reliability of the system above an exogenously determined measure (there is an increasingly expensive tradeoff between reliability and cost)
- One way to calculate system reliability is through the use of a cumulative outage probability table (COPT)
- The COPT generates system states, which are all possible combinations of the generator states in the system, as well as the probability of each.
- It then sums the probability of those events in violation of the reliability criteria (i.e., those states with capacity less than the amount determined via the mapping).
- The number of system states represents all possible combinations of the power plant states whereby the probability of the set of states is determined by the forced outage rate (i.e., the probability of a unit being in unplanned outage).
- The number of states in a two-state model is 2^n and 3^n for a three-state model.
- A 10-unit system has 1024 system states, however, a 30-unit system has over a billion possible states.
- As shown, the COPT table can get quite large even with a small number of power plants, which necessitates alternative methods for determining system reliability.
- One alternative approach are Monte Carlo methods, which use random draws instead of complete enumeration
- But how to incorporate the constraint into the model

$$(1) \quad \text{Reliability}_{\text{MC Sim}} \geq \text{Metric}$$

A class of generation expansion planning (GEP) models approaches the long-term planning of the electricity generation industry using a two-stage framework. In the first stage investment decisions are made (i.e., type, location, quantity). Then in the second stage, random variables are realized, and operation decisions are made (i.e., how much to run each unit, etc.) and constraints, such as the reliability metric are enforced. One issue with the Monte Carlo or COPT approaches, is the lack of information available to the model if the reliability constraints are violated.

For example, the model would first select a mix of generating units in the first stage. Next the model would feed that information, as well as the FOR for each generation unit type to a simulation model. The simulation model would calculate the system level reliability before then feeding that back to the optimization model. The optimization model would check the constraint that the reliability level exceeds the exogenously set metric. However, if the constraint is violated the model does not have the necessary information to select a better set of generation units (i.e.,

there is no linkage between the calculation of this constraint and the first stage investment decisions).

What I'm proposing is to use an approximation method in the first-stage that checks to see that the reliability metric is met before then proceeding to the second-stage. The constraint would use the properties of the generation unit (i.e., size, FOR, etc.) to approximate system level reliability (almost like a chance constraint).

2. POISSON APPROXIMATION

The line of thinking that we have is as follows. The likelihood of a generator being in a particular state (i.e., on or off) follows a Bernoulli distribution. The sum of Bernoulli random variables follows a Poisson distribution. The Poisson distribution is a single parameter distribution defined by its mean λ . Thus, there is a unique mapping between the λ value of a Poisson random variable and, its PDF and CDF.

We can leverage this information to create a constraint that ensures system reliability is above a selected threshold. From the load duration curve we previously estimated the MW capacity associated with the reliability level. We can calculate the λ value of the Poisson distribution that has the unique CDF associated with these values. The λ value also provides meaningful information, it is just the average. Then all we need to do is say that the average FOR across all installed units is greater than the mean value.

However, there is one issue with our approach. The sum of Bernoulli random variables is a binomial random variable if the probability of success is the same for each Bernoulli R.V. If the probability of success is different amongst the Bernoulli R.V.'s than their sum follows a Poisson Binomial R.V., which is generally characterized as having fatter tails as compared to a Poisson R.V. The Poisson Binomial distribution can be approximated by the Poisson distribution with the quality of the approximation depending on the success probability and number of generation units (according to Le Cam's Theorem). But the Poisson Binomial distribution is a two parameter distribution, which means that a unique mapping akin to what we did before does not exist. However, if we can approximate the Poisson binomial distribution using a Poisson distribution then we are back in business.

The Poisson approximation method discussed here approaches the problem of identifying an optimal portfolio of generation assets from the perspective of the targeted portfolio of generators, rather than properties of the individual generation units. Each power plant is modeled as having two states: either on or off with the likelihood of each state determined by the forced outage rate (FOR) parameter a_p . The model selects the number of each power production technology to build p at each node n such that a system of constraints governing the operation of the network are met.

Power system planning involves the determination of an appropriate outage frequency by planners. This is because it becomes prohibitively more expensive to ensure redundancy the higher the probability of success climbs (i.e., the more redundancy built into the system the higher the costs of operating said system). This cost/reliability tradeoff is a central part of any engineered system. In the United States power systems are designed to have system demand exceed system capacity 3 hours a year. With this information, as well as the load duration curve, we can

calculate the λ of the target Poisson distribution. We used the R software package to identify the Poisson distribution whereby the percentile of interest matched with the quantity in the load duration curve. It is important to remember that in the Poisson distribution this is a unique mapping but in the Poisson Binomial distribution, as it is a two parameter distribution, there are an infinite number of mappings.

The constraint to be implemented in the model is that the weighted average available capacity is larger than λ see Equation 2. This constraint leverages information that we necessarily have before we even run the model (i.e., assumptions of future load demand) while also avoiding the need to parameterize reserve capacity into the model. One issue with this approach is the distance between our approximation and the actual Poisson Binomial distribution. I propose that we implement the model and conduct numerical experiments to identify the gap.

$$(2) \quad \sum_{p=1}^P a_p \cdot size_p \cdot \sum_{n=1}^N x_{p,n} \geq \lambda$$