Sizing a Storage System for a Wind Farm
Model A: Statistical Approach
Eric Scott, Jan. 30 2016

Problem

"A wind developer is selling its production directly into the ERCOT day-ahead energy market. Generation deviations greater than 2% from the forecast will be billed based on the real-time energy market price.
Using the forecast and generation data attached, propose the most optimal size of system and describe your approach to model the corresponding dispatch strategy for the battery."

The following write-up explains my step-by-step approach to how I arrive at my final recommendations to the proposed problem.

Step 1: Preliminary Analysis

After downloading the dataset (Wind deviations.xlsx), I wanted to do some very quick pre-analysis checks. I used Excel to get a better understanding of the data before doing any analysis. The dataset has four columns:

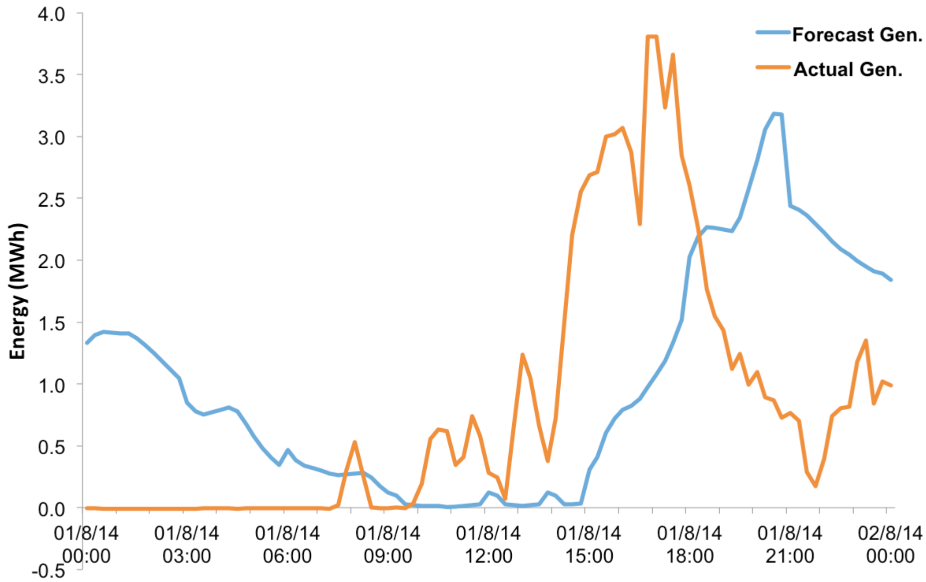
(1) date (day-month-year)
(2) time, 15 min intervals, 24 hr clock (hr:min)
(3) forecasted production (MWh)
(4) actual generation (MWh)

I wanted to get a sense of size for the wind farm. The maximum actual generation was 6.375 MWh. That production is over a 15 minute interval, so I'm approximating the total wind farm potential output as 25.5 MWh. If this is max possible wind production over an hour, assuming 1.5 MW turbines, we have 17 turbines in our wind farm.

Looking a little closer, there were missing values for the time interval. I assumed the rest of the data was continuous in 15 minute intervals and filled in the remainder of the values following the same pattern. I then created a column to combine dates and times into a date/time so that it would be easier (if needed) in python to read as a date-number. Anticipating that I would need to compare the actual and forecasted energy data, I created two more columns in Excel; one for the difference between actual and forecasted energy generation, and one for the deviation percentage measured by:

Deviation = [(actual generation – forecasted generation) / forecasted generation] * 100

Last, I wanted to see what the generation profiles looked like. I plotted the first day in Excel (again as a sanity check):



The plot shows considerable variation between forecasted and actual generation for August 1, 2014. The modified Excel data was saved and imported into python for analysis.

Step 2a: Describing the Data

I used python for the rest of the analysis. In python I used the pandas DataFrame and Series functionality to store and manipulate the data. Using the *describe* function, we see that:

	Forecasted (MWh)	Actual (MWh)
Mean	1.66	1.56
Standard deviation	1.49	1.71
25% quartile	0.25	0.002
50% quartile	1.37	0.99
75% quartile	2.79	-2.65
Minimum value	0	-0.13
Maximum value	6.3	6.36

From the table above, the actual production tends to be lower than forecasted. The standard deviation is also quite a bit higher in the actual generation than forecasted. With the same approach for our 'delta' column (actual generation – forecasted generation) we see that the mean delta is around -10.36 MWh. We run into some problems when trying to do this on the deviation data.

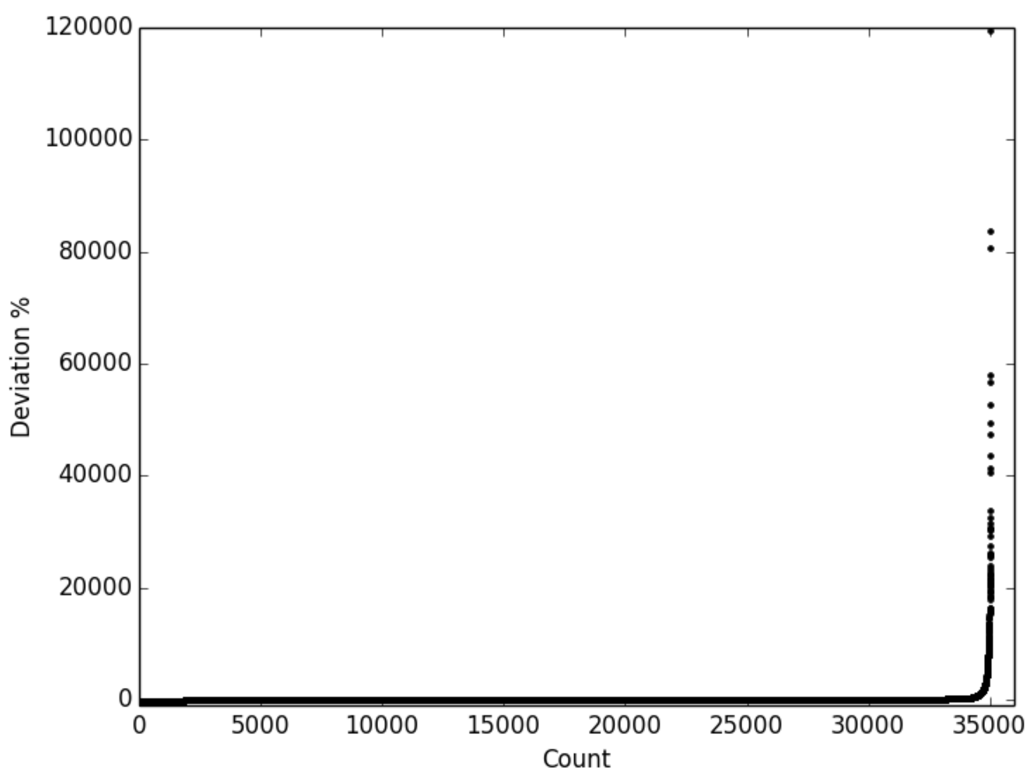
Step 2b: Cleaning the Data:

Right off the bat, we cannot calculate any summary statistics on the deviation column because there are non-numerical entries. When we calculated the deviation before, there are certain cases where Excel input a value of 'DIV0!'. There are also values for +/- infinity, and NaN. I set values for +inf, DIV0! and NaN to 2%, and values for –inf to -2%. This would be enough to trigger the condition of selling into the real-time market. These values occurred primarily when we forecasted no generation, but actually generated energy from out wind farm.

When we look at some simple statistics on the newly modified deviation column, we see that it is skewed by outliers:

	Deviation (%)
Mean	75.53
Standard deviation	1526.34
25% quartile	-79.99
50% quartile	-14.66
75% quartile	-15.97
Minimum value	-587.67
Maximum value	119466.00

These numbers don't agree with the actual and forecast data. The mean deviation is a positive number, however we know from the previous table that we actually produce less than forecast and therefore would expect a negative deviation mean. We do not want to see such a large value for the standard deviation, as well as the very high max and min value. Plotting a sorted distribution we can see the outliers:

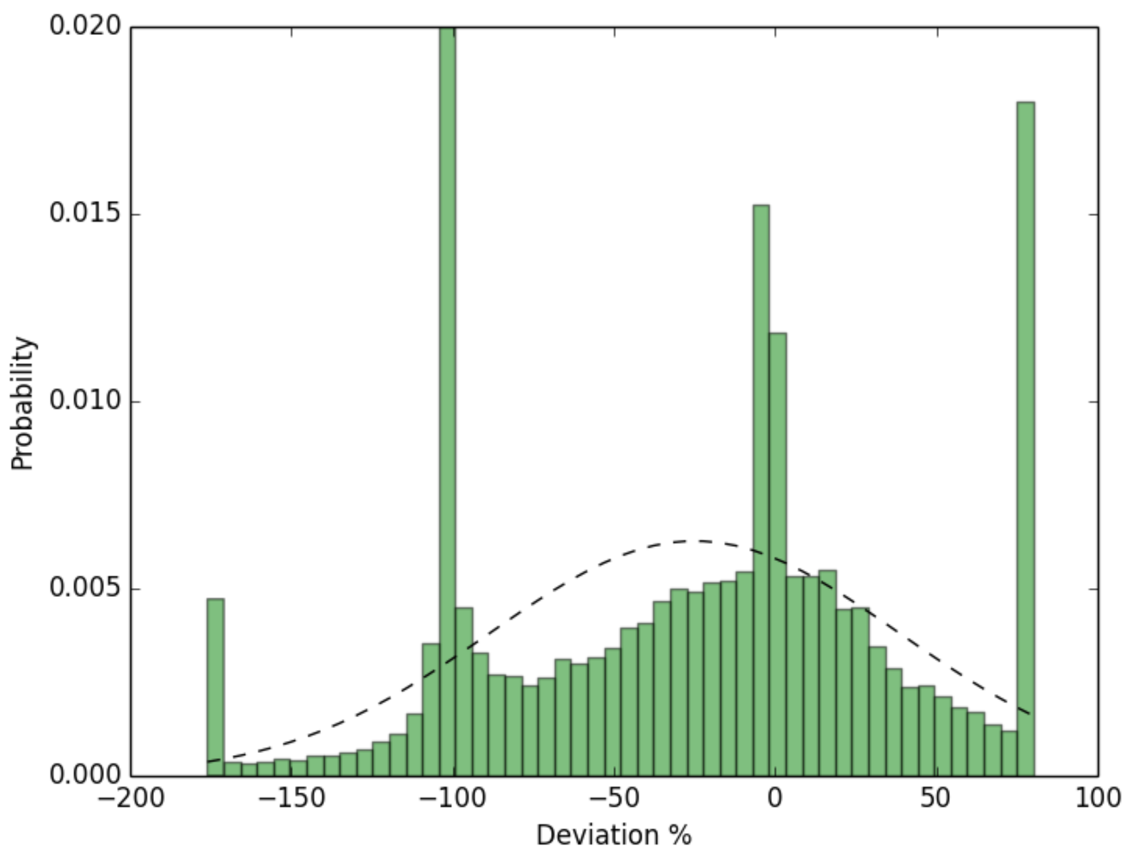


A couple data points that have extremely large deviation are skewing the data. Again this is occurring when we forecast no production but end up generating energy. Doing a little investigation, we see the max deviation occurs on Sept. 4 at 10:14am. We forecasted 0.0025 MWh of delivery (essentially nothing), and actually generated 2.98664 MWhr!

All of this is important because I will use a rudimentary statistical approach to size the battery system – so I need to ensure the data makes sense. The high and low outlier limits are defined by:

High Limit = Q1 – IQR*1.5
Low Limit = Q3 + IQR*1.5

Where the IQR is the interquartile range, and is Q3-Q1 (Q3 is the third quartile and Q1 is the first quartile). Our new upper and lower limit are 80.06% and -176.02%. I set all values greater or lower than the limits equal to the limits themselves. With the modified deviation we have a new distribution shown with the probability density function (PDF):



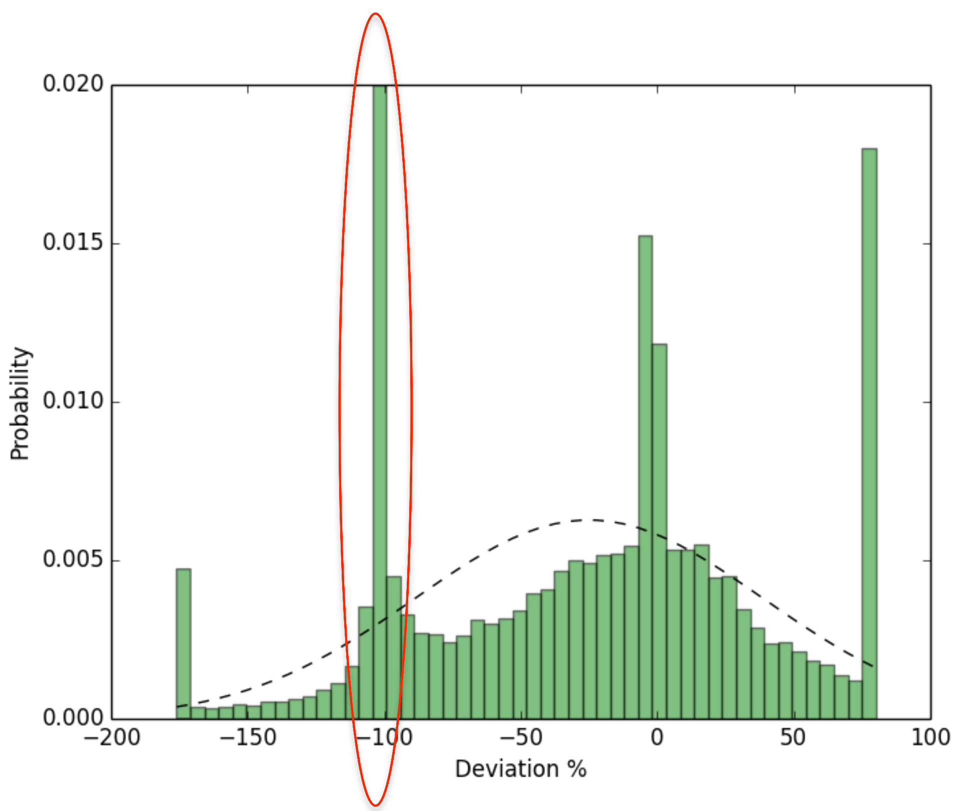
We see 4 spikes that deviate from the PDF. The lowest spike occurs were we set all values from the outliers at the lower outlier limit (-176%). We see the largest spike occurring around -100%. This is happening when we forecast generation, but actually generate almost no energy. We see a spike around + and – 2% because that's where we set non-numeric values. Finally we see the spike at the upper outlier limit for similar reasons to the lower outlier limit.

From the histogram it looks like the vast majority of data fall outside that 2% deviation. Doing a little more investigating I found that about 89.5% (31375 of 35040 data points) fall outside the +/- 2% deviation.

Step 3: Sizing the Battery

The way our problem is set up, we sell energy generated if within 2% of forecasted, and the rest is 'billed' at the real-time rate. There are two scenarios here: first is that we over-generate and receive payment (or credit) for energy in excess of the 2%, second is that we under-generate and are charged (or debited) for energy that was promised but not delivered.

Energy markets are very complex and have numerous pricing schemes. For the sake of this model I must consider peak and off-peak times. My main goal for the developer is to **minimize the amount charged during peak times.** If the developer promises energy during peak pricing times but does not deliver, they are putting themselves through extreme financial burden. We have seen so far two things that are cause for concern in the previous steps. First is the large difference between actual and forecasted statistics. This already hints at the possibility of some nasty charges in the real-time markets. The second confirms this suspicion: the spike in the PDF at -100%, which means that we are under-delivering on many occasions:

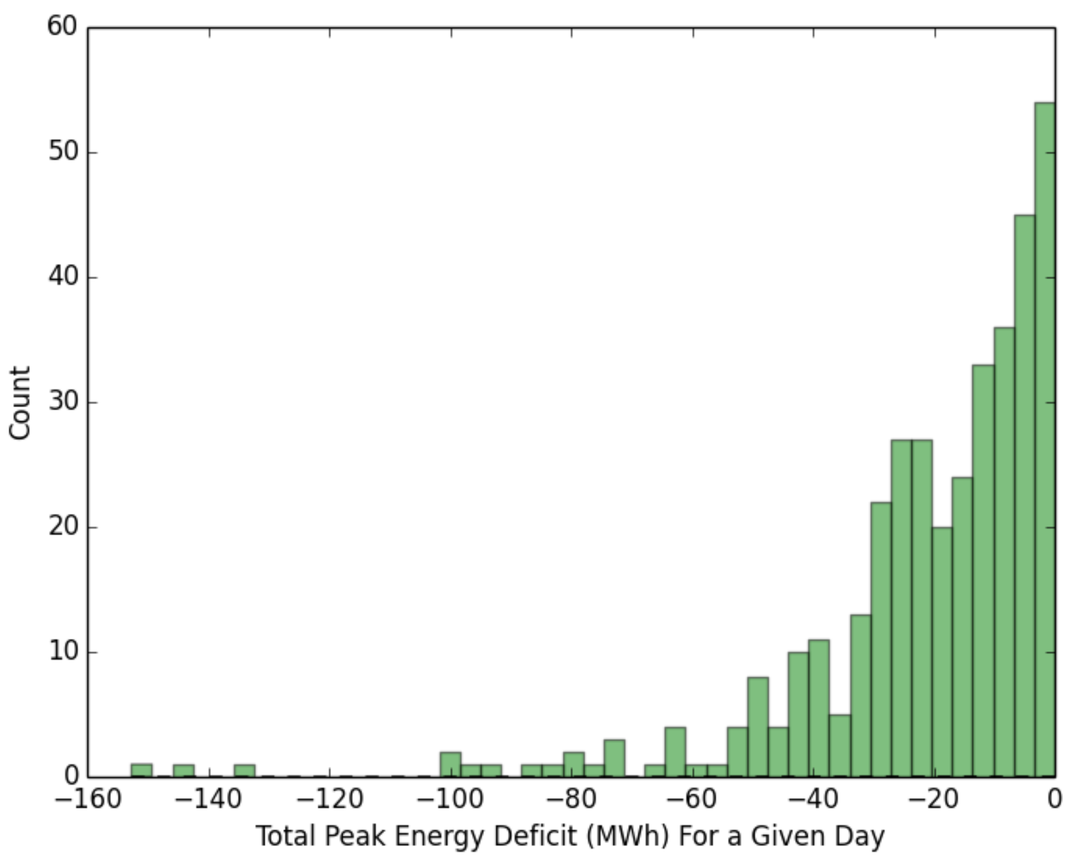


Again, we want to minimize this cost to the developer by using a battery storage system. ERCOT states that peak hours are 07:00-22:00 Monday to Friday. For this analysis, I assume all days Monday through Sunday between 07:00-22:00 are peaks times. All other times are off-peak. I also ignore holidays.

I wrote a script in python that calculates the *sum of peak-charges each day.* My script sums the charges that occur during peak-pricing period for every day in our data set, and ignores both non-peak values and surplus energy sold. Running some summary statistics on this:

	Deficit Charge (MWh)
Mean	-21.29
Standard deviation	21.82
25% quartile	-28.39
50% quartile	-15.91
75% quartile	-6.207
Minimum value	-152.91
Maximum value	0

The above table shows that of all days in the dataset, we under-produce (and are charged for) an average of 21.29 MWh. A histogram helps visualize the distribution:



The distribution stops at 0 because we only count the deficits (negative deviations). I will size my battery storage system to land inside the 50% quartile, that is to have the capacity to meet 50% of all the day's energy deficits. I do not use the mean because I believe this would over-size the storage size due to anomaly days. **Based on this, my initial choice for a battery storage system capacity is 16 MWh.**

Step 4: ERCOT Assumptions

Before coming up with the battery storage dispatch strategy, I needed to make several assumptions about the energy market. As mentioned above, energy markets are very complex. There are numerous ways the market must balance supply and demand. My assumptions include:

- Real-time market charges are billed at the LMP (locational margin price). The LMP is assumed to be fixed at **$200/MWh during peak** times and **$20/MWh during off peak** times. (These values were approximated based on www.energyonline.com and www.ercot.com). In reality the LMP varies a lot depending on grid conditions.
- Related to the above assumption, we are assuming only positive LMP pricing, in reality this goes negative under certain grid conditions.
- Actual energy generation that falls within 2% (+ or -) of forecasted is sold at the day-ahead (DA) rate. For this analysis, I assumed the **DA rate is fixed at $20/MWh**. In reality this would change every day.
- Peak hours occur from the 7:15 measurement to 22:00 on all days of the week. Off-peak are all other times (0:00 – 7:00 and 22:15-23:45). In reality peak times might not include weekends and holidays.
- There are no breach of contracts and no taxes on transactions, no other expenses are considered (operational, etc.).
- There are no grid outages.
- There are no power balance or storage constraints on our battery system.
- We do not sell ancillary services from our storage system.
- Markets are settled at either the LMP, DA or both at the end of every 15 minute interval. In reality the markets are settled at settlement prices after each operating day.

Assuming the LMP (peak and off-peak) and DA pricing are fixed is a huge assumption, but helps illustrate the dispatch strategy.

Step 5a: Dispatch Strategy – No Storage

Let's start with our initial condition, just the wind farm and no battery storage. There are 3 scenarios:

Scenario 1: If our energy deviation is **LESS** than 2% from forecast, we sell actual energy produced at the DA rate.

Scenario 2A: If our energy deviation is **GREATER** than 2% from forecast and we have **OVER-GENERATED**, we sell forecast energy produced at the DA rate and sell our excess at the LMP (peak or off-peak depending on the time).

Scenario 2B: If our energy deviation is **GREATER** than 2% from forecast and we have **UNDER-GENERATED**, we sell actual energy produced at the DA rate and pay the difference to what we had committed at the LMP (peak or off-peak depending on the time).

This is not a dispatch strategy but more of a grid operation rulebook. These are the rules we must follow if want to be a wholesale electricity provider. It also helps provide a framework for what to do when we have battery storage. Using the values mentioned in the assumptions, our wind farm generates **$392,144** in income over the year. Running a quick analysis, it turns out that we paid ***$1,292,514*** over the year in the real-time peak LMP due to *under-generation!* That is over a million dollars in lost revenue due to over-commitment. This is why the battery storage system is designed to alleviate the financial burden of under-generation during peak times!

Step 4b: Don't Go Too Far Down the Rabbit Hole!

Before jumping into the battery dispatch strategy, I wanted to see the sensitivity the initial model had to the assumed prices. I started to build a script that would produce a 3D map with price variation. I would vary the peak and off-peak LMP prices to test - among other things - negative LMP and extremely high LMP rates. After running a couple scripts, the computation time started to get very long. Each model for a given peak and off-peak LMP cycles through about 35,000 values. Testing 5 values each for the peak and off-peak LMP prices adds 25 cycles of 35,000 values each – which is 875,000 values to go through.

I started to see how complex and involved the analysis could get. It also dawned on me that there is historical pricing available online (not for ERCOT for more than 30 days but for other energy markets). I decided not to take my price sensitivity analysis too far but just come up with some general ideal about price sensitivity. Running the model under various prices, the wind farm income could vary from $70,000 up to $500,000 – increasing when we increase both peak and off-peak LMP.

Step 4c: Battery Dispatch Strategy

Using the strategy mentioned in 4a, I added some conditions to include battery storage. We now care about peak and off-peak times:

00:00 – 7:00 or 22:15 – 23:45 (OFF-PEAK) ***Rule: store excess if capacity available***	**Scenario 1a:** If the deviation is less than 2%, sell at the DA rate. **Scenario 1b-1:** If deviation is **GREATER** than 2% and we **OVER-GENERATE** and there is **AVAILABLE** capacity in the battery, sell the forecast generation at the DA rate and store the excess in the battery. **Scenario 1b-2:** If deviation is **GREATER** than 2% and we **OVER-GENERATE** but there is **LIMITED** capacity in the battery, sell the forecast generation at the DA rate, store until the capacity is met and sell the remainder at off-peak LMP. **Scenario 1c:** If deviation is **GREATER** than 2% and we **UNDER-GENERATE**, sell our actual generation and pay the off-peak LMP price.
7:15 – 22:00 (PEAK) ***Rule: use storage to avoid LMP if available***	**Scenario 1a:** If the deviation is less than 2%, sell at the DA rate. **Scenario 1b-1:** If deviation is **GREATER** than 2% and we **UNDER-GENERATE** but there is **AVAILABLE** energy stored in the battery, sell the actual generation at the DA rate and use stored energy to avoid LMP. **Scenario 1b-2:** If deviation is **GREATER** than 2% and we **UNDER-GENERATE** and there is **LIMITED** stored energy in the battery, sell the actual generation at the DA rate, use all available energy in the battery to offset LMP and pay the remainder at peak LMP. **Scenario 1c:** If deviation is **GREATER** than 2% and we **OVER-GENERATE**, sell our forecasted generation at the DA and sell the excess at the peak LMP.

The above dispatch strategy can be simplified by the following: if during peak hours, use the battery to avoid paying LMP. It if's off-peak, store excess energy in the battery if there's space. We want to avoid paying high prices for under-generation, and want to avoid getting paid low prices for our over-production.

Running this simulation in python, the dispatch strategy with a capacity of 16 MWh increases net income to **$3,045,665**, an increase in more than **675%!**

Final Words

This is an initial approach to solving the proposed storage problem. It provides a good starting point for sizing a battery storage system based on statistical deviations in actual and forecasted generation. A wind farm's production is highly variable (unlike a more stable coal and natural gas power plant). The **key to the dispatch strategy** is that the battery allows the wind farm developer to avoid the peak LMP prices.

There are numerous other factors that should be analyzed and included in the model. This did not provide a good framework for financial payback, or look at any technical considerations of the battery system. This model does not look at power capacity and rates of charge and discharge. It has huge areas for improvement and increased complexity – for that I moved to Model B.