Hybrid Renewable Energy Capacities for Target Grid Penetration

Ankita Deshmukh, Akhilesh Ramakrishnan, Kaustubh Upadhyay
Departments of Mechanical Engineering and Computer Science
Columbia University
ad3293@columbia.edu, ar3539@columbia.edu, ku2151@columbia.edu

Abstract— **In this project we seek to use big data analytics platforms to compute the optimal capacity of renewable electricity generation resources to be installed in order to supply a set target fraction of the total energy demand. We optimize the combination of resources based on capital cost under the constraints of user-set values of energy wastage and grid penetration. Optimization is done using the genetic algorithm. This report will focus on the performance and results of this software when input with datasets for New York state for the year 2010. However, it is capable of working with any hourly load profile, solar irradiation, and mean wind speed data.**

Keywords-renewable energy; optimal sizing; hadoop; genetic algorithm; New York State; wind energy; solar energy; grid integration

I. Introduction

The world is increasingly looking to renewable energy resources for energy needs due to a variety of reasons. Air pollution, environmental effects, depletion of natural resources, and cost are all driving factors in this gradual change. One of the problems with solar and wind energy in particular is their intermittent nature. Energy output from both these sources vary with the weather conditions.

Every state in the USA has a Renewable Portfolio Standard (RPS) where they commit to meet a certain specified portion of the annual energy demand through renewable energy. They must also set targets to reach a certain level of grid penetration with renewables by a certain target date. Grid penetration refers to the percentage of the annual total energy demand that is met by a certain type of generation resource.

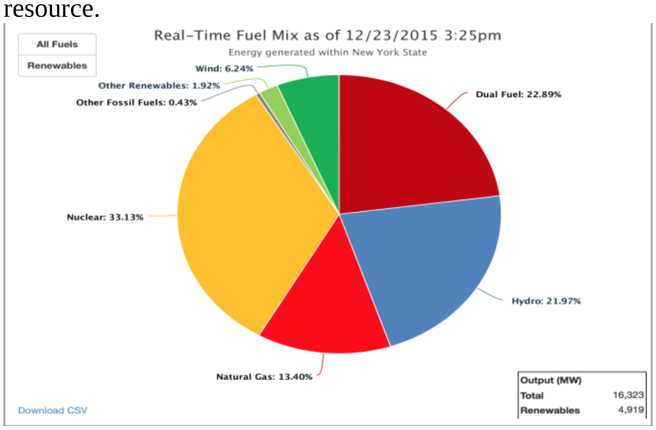

Figure 1 Energy Mix of New York State - [Source]()

The state of New York has set a target of 50% grid penetration with renewables by 2030. Its current energy mix is shown in figure 1. Currently about 30% of the demand is bring met by renewable energy in which most of the contribution is from hydroelectric power plants. Over the next decade, as many of the ageing nuclear reactors come offline and the state seeks to reduce fossil fuel use, the demand must be met by increased use of renewables. This will mainly have to be wind and solar energy as the hydroelectric plants require huge infrastructure projects and most of the rivers in the state have already been tapped.

This poses the natural question of what capacity of renewable energy generators the state must install to meet its target. This problem is complicated by the fact that each state has different load profiles. The load profile of each state also varies through the seasons, mostly due to the change in heating and cooling loads. Peak demands occur at different times of the day and vary in amplitude. This implies that the analysis must be done at least at the resolution of an hour in order to accurately determine parameters such as reliability, energy wastage, and grid penetration.

In general, it can be seen that the wind speeds peak during the early mornings and evenings, whereas solar insolation peaks during midday. The load, with some variations, usually peaks at some point between noon and sunset.

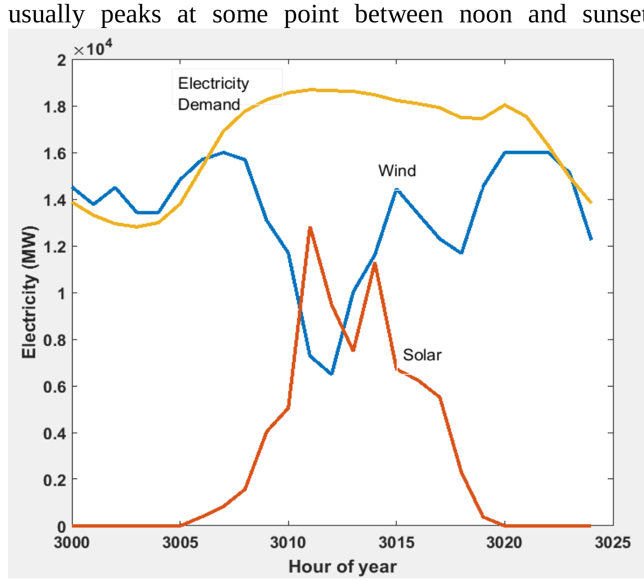

Figure 2 NY State Load vs. Power supply from 30 GW Capacities for a day in April

As seen in Figure 2, wind energy is low during the period when the demand is high and solar energy is zero apart from a few hours during the day. Using just one of these forms of energy would lead to an energy shortage at one point of the day and an energy excess at another point of the day. This problem is mitigated by combining both in order to achieve a curve more similar to the demand curve.

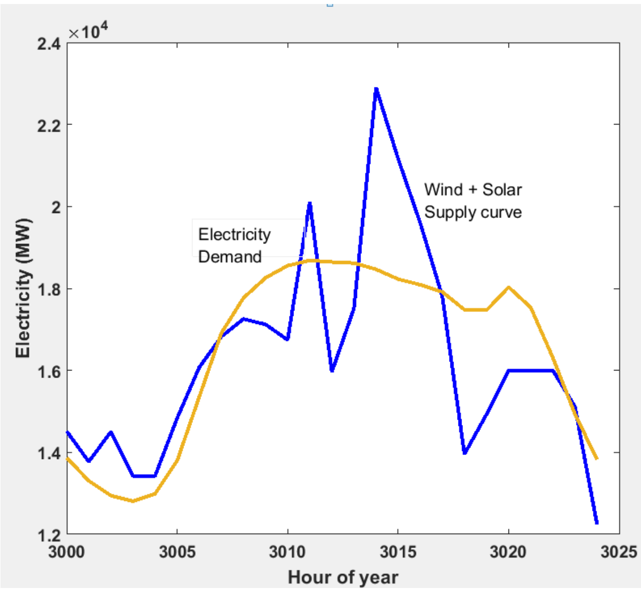

Figure 3 NY State load vs. Combined Solar & Wind Supply for a day in April

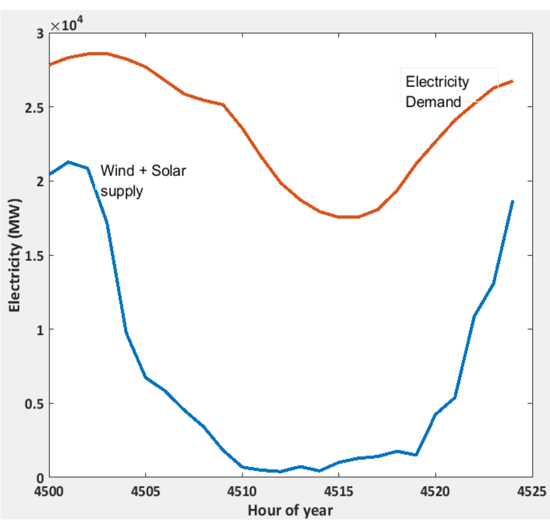

Figure 4 NY State load vs. Combined Solar & Wind Supply for a day in July

From Figures 3 and 4, we can see that there is a large variation in the supply and demand curves between the seasons. Clearly, a combination of the two resources is required to minimize energy wastage and improve reliability, on a daily basis. However, grid penetration refers to the percentage of *annual* energy met by a certain resource. This means that a grid penetration target can be met even if during certain days or months of the year, the supply from these resources is very low or even zero. In other words, reliability is not an issue. However, energy wastage is an issue, as the owner of the generation asset will want to maximize income.

These considerations led us to set the main objective function to represent the capital cost. Cost optimization is done under the constraints of energy wastage and grid penetration. In other words, the output from this will be a certain combination of solar and wind energy capacities, which will meet a given percentage of the annual energy demand while keeping energy wastage below a certain percentage of the generated energy. The two constraints may be specified by the user depending on the target penetration and cost of energy.

II. Related Works

In “The technical, geographical, and economic feasibility for solar energy to supply the energy needs of the US”[2], Fthenakis et al., several scenarios over 5-year time periods up to 2050. They consider the problems that would have to be addressed as the penetration of solar energy into the grid increases even up to 90%. They also consider the impact of combining different types of energy resources into the energy mix. This paper makes a case for reduction of technology costs, improvement of energy storage options, and combined use of different types of renewables.

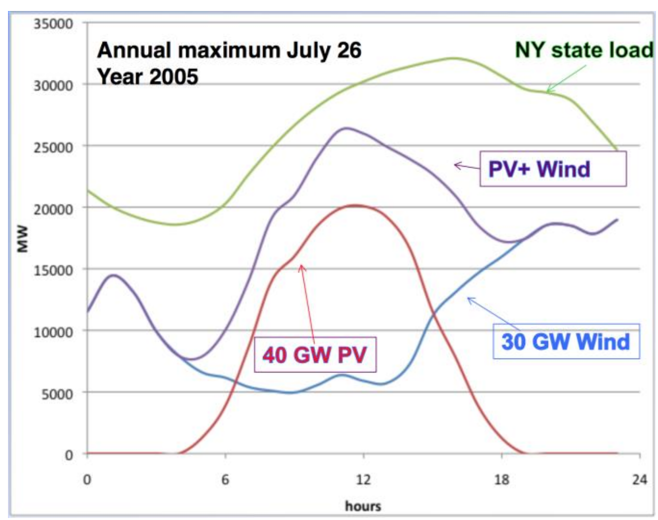

Figure 5 NYS 2005 Load vs. PV and Wind Resources [1]

Figure 5 [1] makes a case for combining wind and solar energy in order to match the New York State load. Even though solar insolation in the state is below average, it helps improve the supply curve. These ideas are integrated in our approach to optimal sizing.

In "The Optimal Sizing of Standalone Hybrid Wind/PV Systems Using Genetic Algorithms"[3], Xu et al. provide a method to find the optimal size of a standalone hybrid system. They include the use of batteries and optimize for cost and reliability. We follow a similar approach to the problem but ignore reliability and consider the two constraints mentioned in the previous section instead of reliability. Also, the grid level problem involves huge datasets as opposed to the standalone system level problem which is much less data intensive.

III. System Overview

The system takes in three .csv files as input. The following is the list of input files required for the program:

1. Hourly load data for 1 year (megawatts)
2. Hourly solar power data for 1 year (megawatts)
3. Hourly wind power data for 1 year (megawatts)

The datasets used were:

1. NYISO real time load data: Source - The NYISO provides real time load data and has archives for many previous years. This is given in terms of zones. This data was pre-processed in order to combine the load of all the zones and shown one value of load for the entire state for each hour.
2. Solar insolation: "The Solar Prospector", NREL - Source. This gives the hourly solar insolation values in terms of kWh/m^2. This data was preprocessed to give the hourly power output from a 1MW solar array fixed at the latitude tilt.
3. Wind Speed: "Eastern Wind Resources Dataset", NREL - Source. This gives the mean wind speed at a resolution of 5 minutes at a height of 100 meters. This data was preprocessed to give the hourly power output from a 5MW wind turbine.

The user must specify the constraint values, ie. The minimum percentage value of annual energy that must be supplied and the maximum percentage value of energy that may be wasted.

The output is a .csv file which in the form of an array with columns of the form [Number of 1MW Solar Arrays| Number of 5MW Wind Turbines | Total Capacity Installed (kW) | Total Capital Cost ($)]. This file contains all the considered combinations that satisfied the constraints. The last entry is the optimal combination. If there is no possible combination that satisfies the input constraints, then this file will contain only one entry, the combination that gets closest to satisfying both the constraints.

IV. Algorithm

The problem of optimal sizing of a hybrid renewable power plant is in essence a problem of multimodal optimization. However the relationship between the various parameters that are being used to size said plant can often get complicated owing to the inherent non-linearity in their relation (e.g. Cost of the plant is affected by amongst other things the resale value of the components which is affected by amongst other things the annual interest rates) making it difficult to solve such a problem using conventional optimization methods. In such a situation when one is constrained by computational resources one is best served by a heuristic based optimization approach. Genetic algorithms (GA) and Particle Swarm Optimization (PSO) algorithms are the two widely used solutions to this problem. Both work on the basis of population of potential solutions and both deliver a comparable quality of results [4] however as it stands GA is an inherently discrete technique (PSO assumes a continuous domain) as well as being faster and overall easier to implement as compared to PSO.

Genetic Algorithms belong to the more general class of Evolutionary Algorithms and at its essence this algorithm mimics the process of natural selection and employs the principle of "survival of the fittest". They were introduced in the mid-70s by Dr. John Holland [5] at the University of Michigan. It starts with a population of potential solutions called chromosomes or phenotypes. These phenotypes constitute the first generation. Each of these phenotypes is then evaluated for fitness using a fitness function. The fitness score so obtained reflects the extent of objective function minimization as well the as the satisfaction of the constraints by the phenotypes. Using a procedure such as s-wise tournament the fittest phenotypes of a generation are selected to undergo a procedure called crossover. Just as genes share information during reproduction so do the fittest phenotypes so as to ensure the propagation of the 'fittest' features from the current generation to future generation. This also ensures that fitness of phenotypes increases across successive generations. However, it can so happen that all the initial phenotypes belong to one portion or neighborhood of the search space while the minimum lies is another. There must be, therefore, a mechanism to introduce 'diversity' randomly in the population. This is provided by the Mutation operation where in at random a phenotype modifies its own features. This process is then repeated across a number of generations until one or more of the many stopping criteria are met e.g. fitness threshold, time threshold, generation threshold etc. This process works because of Holland's schema theorem [6] which states that short low order schemas of above average fitness increase exponentially across generations.

In this project GA has been implemented using the genetic algorithm toolbox in MATLAB. It follows all the steps listed above. Selection is done using scaled fitness scores using a process called 'Stochastic Uniform' which lays out a line in which each parent corresponds to a section of the line of length proportional to its scaled value. The algorithm moves along the line in steps of equal size. At each step, the algorithm allocates a parent from the section it lands on. It uses uniform crossover (any and all attributes can be crossed randomly) and implements mutation by adding a Gaussian noise vector to the representation or phenotype. Stopping criteria can be specified, here it was fitness limit.

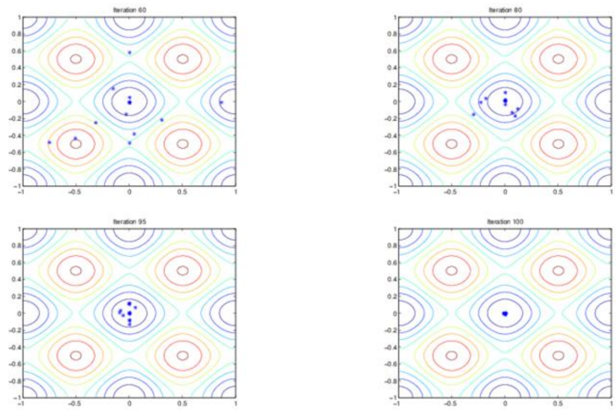

Figure 6 Plot of how the initial population converges to a solution across generations [1]

GA thus provides a method to obtain a 'good enough' solution in a 'reasonable amount' of time. However, there are some problems with this approach. Firstly, there is no guarantee that it will converge to the global minimum or even (depending on the stopping criteria) to a local minimum. The quality of the final solution is relative to the other members of the populace. Also, the number of generations needed to converge to a reasonable solution goes up exponentially as the number of parameters in the representation of the phenotype go up.

In a plant that costs several billion dollars to construct any further optimization can save millions of dollars if not more. The Mapreduce algorithm provides a way to execute an operation in parallel by splitting it into a map stage wherein the input values, in the form of key-value pairs, are mapped to an intermediate set of key-value pairs and a reduce stage wherein these intermediate value are combined or reduced key-wise. Apache Hadoop is an open source software library which provides a framework to implement mapreduce on not just one but several nodes. The authors believe that the distributed processing power provided by Hadoop combined with the meta-heuristic approach of GA is the best way to solve the problem at hand.

Consequently, the following scheme is advised. GA is run multiple times and the mean of the solutions is determined. Based on this mean an arbitrary neighborhood is defined and a file is generated containing all the points in that neighborhood.

A mapreduce job is constructed using the Hadoop libraries wherein the map part consists of calculating the objective function and the constrained parameters for each possible input. The reduce part consists of finding the intermediate value with lowest value for the objective function. This file is then compiled and compressed in a jar. The input file is transferred to the Hadoop distributed file system, YARN is started and then using this jar the mapreduce job is submitted to Hadoop. It makes sense to apply 'Brute Force' in the neighborhood of the solution suggested by GA as has been discussed then tend to converge towards the optimal solution rather than directly on it. Applying GA using mapreduce would not have gotten rid of the inherent problems in GA that have been discussed above. At the same time unless one has a huge cluster of several nodes it is not conceivable to brute force over all the points in order to find the optimum solution. The scheme discussed above develops on the strengths of GA while addressing some of its weaknesses and being computationally feasible on a small cluster during that entire time.

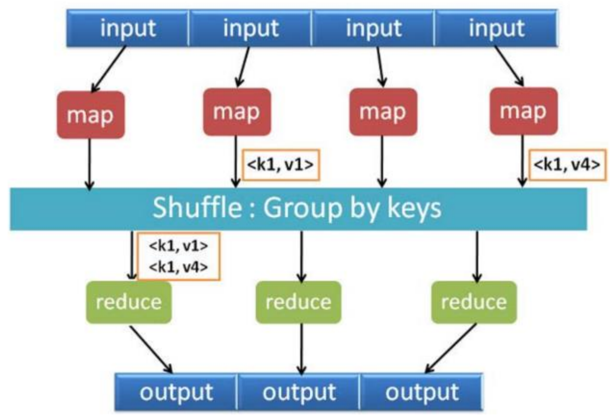

Figure 7 Schematic diagram of MapReduce (from the Apache Software foundation)

V. SOFTWARE PACKAGE DESCRIPTION

We have used the genetic algorithm toolbox in MATLAB (student license) to implement the actual heuristic optimization and we have used Apache Hadoop 2.7.1 to refine it. Jobs are submitted to Hadoop using *.jar files. The main files are as follows. On the MatLab side we have "Call.m", this contains the actual call to the genetic algorithm. "Cost.m" and "Constraint.m" contain the objective and constraint functions respectively, these functions are passed to the genetic algorithm so that it can call it during the fitness evaluation of each generation. BruteForceWriter.java when given a range generates a csv file containing all possible input combinations. This is the input file and is uploaded to the HDFS. Optimizer.java contains the actual mapreduce job, this file is compiled and its jar is submitted to Hadoop.

VI. EXPERIMENT RESULTS

A. Optimal solution

The optimization was run on the data of year 2012 for the state of New York. To maintain simplicity, the demand data for all zones in the state were added to give a single demand curve. This method assumes the uniform grid connectivity for all the zones. The following plot shows the hourly electricity demand for the year. As we notice, the annual demand peaks in the summer. This can be attributed to the longer day durations in summer and the fact that in New York, some of the cooling loads are served by electricity driven air-conditioning while the heating loads are majorly served by Natural gas boilers. The peak demand of NY state is 33.86 GW and the annual total energy demand is 16335 GWh.

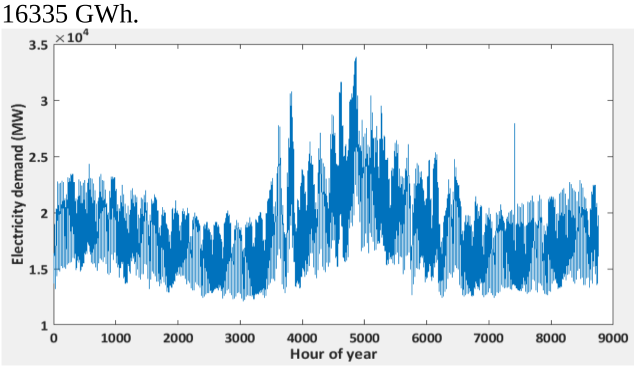

Figure 8 Annual electricity demand for NY

As the problem statement denotes, the purpose is to determine the right combination of the solar and wind electricity capacity to meet this demand. Following Figure 2 shows the supply curve from the installation of 5MW of wind and 5MW of solar.

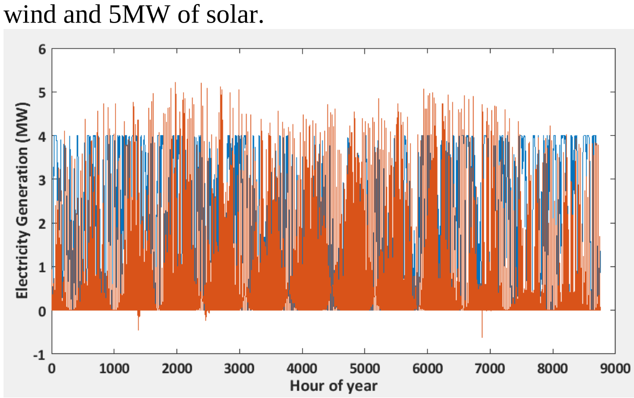

Figure 9 Electricity supply plot with 5MW solar & wind

The cost optimal solution to meet the annual electricity demand and maintain a certain level of wastage in the system is obtained using genetic algorithm optimization. The model was run for several combinations of the levels of energy coming from renewable resources and levels of the wastage in the system. Following table shows the solution for meeting 50% of energy demand and excess energy level of 20%. This gave the solution with 224 MW of solar and 30510 MW of wind power installations at the minimum cost of 40 BN $. After optimization the solution gave only 14% of excess. Wind capacity is very close to the peak demand of the state.

Table 1: Optimal cost solution for 50% energy demand & 20% excess energy level

Solar Capacity	224 MW
Wind Capacity	30510 MW
Capital Cost	$ 40 BN
Excess Energy %	14.24 %
Grid Penetration %	50%

This optimization ran for 8042 iterations out of which 5774 points met the constraints and ran for ~25 minutes. Figure 3 shows the supply and demand with the optimal solution. As we see, the supply curve from renewables has no pattern and is random in nature. 224 MW solar is utilized at capacity factor of 16.24% while 30.5 GW of wind is utilized at the capacity factor of 35.51%. These numbers are similar to the observed capacity factors for solar and wind.

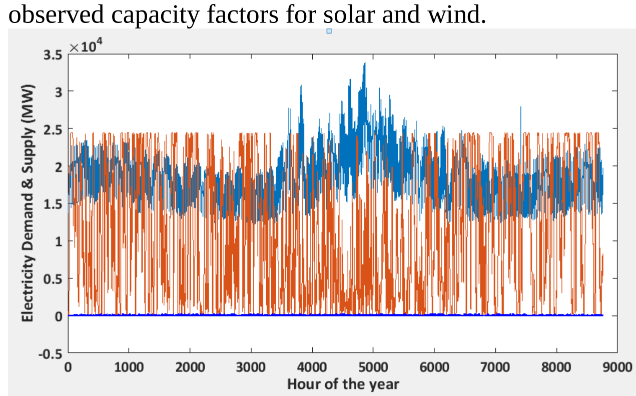

Figure 10 Optimized supply curve for 50% and 20% wastage

B. Different Combinations

Optimization model was run for different combinations of energy demand and wastage. For NY State wind speeds are relatively high whereas solar insolation. This means for the same capacity installed, wind resources will produce more power than solar and hence the cost of energy from win resources is much lower. This is clearly reflected in the optimal combinations shown below in Table 2 where wind resources are overwhelmingly favored. It is only at very low levels of permitted energy wastage that we see large capacities of solar energy required (40% penetration & 0% wastage, 50% penetration and 5% wastage). This is due to the reasons mentioned in previous sections where we explained that the supply curve matches the demand more closely when both solar and wind energy are used. This is because solar supply peaks at the same time as the peak of the electricity demand i.e. during the middle of the day. On the other hand, wind power peaks during the morning and nights. Hence, to minimize the wastage in the system, the optimization prefers higher solar power installation as compared to wind. The 50% penetration column clearly shows how the capacity of solar energy required increases as the allowed percentage of energy wastage is reduced.

Table 2: Optimal solutions for different combinations

Grid Penetration	10%	20%	30%	40%	50%
Wasted Energy	**Capacities – Solar and Wind (GW)**				
0%	0 5.3	0 10.5	0 15.7	8.2 17.3	-
5%	0 5.3	0 10.5	0 15.7	0 21.5	10.5 22.8
10%	0 5.3	0 10.5	0 15.7	0 21.5	3.75 27.5
20%	0 5.3	0 10.5	0 15.7	0 21.5	0.22 30.5

Another interesting result is that at 50% penetration, it is impossible to have 0% wastage of energy. This shows that even when combining solar and wind resources, the variation between seasons is too much to have a certain capacity combination match the load curve perfectly.

The cost of each combination is not shown in the table as they can be easily interpolated from the rates used in the calculation. Solar power was take to cost \$1500/kW$_p$ and wind \$1300/kW$_p$. The main aim was to find the optimal combination for 50% penetration and 20% allowed wastage and that case is discussed in previous sections.

VII. CONCLUSION

This project provides a reliable method for policy makers and planners to get an idea of how much it would cost to gradually move the grid from fossil fuel dependency to renewable energy resources. It is clear that some form of controlled variable generation will still be required to exactly match the demand. Renewable resources can be used to meet any percentage of the demand given that energy wastage is not a constraint. Fast ramping generators and spinning reserves will still be required to ensure reliability and to meet any contingencies. However, we can state that the cost figure of \$40 billion mentioned above is realistic given that it is the capital cost for completely new generation facilities to be installed. This figure must be taken in the context that it will be spread out over a couple of decades and is comparable to conventional energy costs.

ACKNOWLEDGMENT

THE AUTHORS WOULD LIKE TO THANK PROFESSOR CHING-YUNG LIN AND ALL THE TEACHING ASSISTANTS FOR THEIR VALUABLE TIME AND GUIDANCE OVER THE COURSE OF THIS PROJECT.

REFERENCES

[1] Nikolakakis, Thomas, and Vasilis Fthenakis. "The optimum mix of electricity from wind-and solar-sources in conventional power systems: Evaluating the case for New York State." Energy Policy 39.11 (2011): 6972-6980.

[2] Fthenakis, Vasilis, James E. Mason, and Ken Zweibel. "The technical, geographical, and economic feasibility for solar energy to supply the energy needs of the US." Energy Policy 37.2 (2009): 387-399.

[3] Xu, Daming, et al. "Optimal sizing of standalone hybrid wind/PV power systems using genetic algorithms." Electrical and Computer Engineering, 2005. Canadian Conference on. IEEE, 2005.

[4] R. Hassan, B. Cohanim, O. De Weck, and G. Venter, "A comparison of particle swarm optimization and the genetic algorithm," in *Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference*, 2005, pp. 1–13.

[5] Goldberg, D., Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989, pp. 1-25.

[6] Bridges, Clayton L.; Goldberg, David E. (1987). *An analysis of reproduction and crossover in a binary-coded genetic algorithm*. 2nd Int'l Conf. on Genetic Algorithms and their applications

[7] https://www.mathworks.com/help/gads/how-the-genetic-algorithm-works.html