Daily Market Analysis Capability and Results

Nuclear Fuel Cycle and Supply Chain

Prepared for

U.S. Department of Energy

Systems Analysis and Integration Campaign

N. Stauff, G. Maronati, R. Ponciroli, F. Ganda, T. Kim, T. Taiwo (ANL), A. Cuadra, M. Todosow (BNL), P. Talbot, C. Rabiti, B. Dixon (INL), S. Kim (PNNL)

April 30, 2019

ANL/NSE-19/5

EXECUTIVE SUMMARY

Nuclear energy plays an important role in the U.S. energy mix that will likely need to be maintained or strengthened to achieve significant greenhouse gas reduction. However, maintaining the nuclear portfolio becomes increasingly challenging in the current U.S. energy market since the low price of natural gas and the penetration of subsidized and low-marginal cost variable renewable electricity (VRE) are affecting the profitability of nuclear units. In this context, building new nuclear power plants will require increased competitiveness with reduced capital and O&M costs and increased revenues enabled by changes in market policies or increased flexible operation. Within the U.S. Department of Energy, Office of Nuclear Energy, the System Analysis and Integration Campaign has been acquiring the capability to model energy market economics in order to assist decision makers and nuclear utilities. The methods developed and codes acquired are displayed in Figure 1. The objective of this report is to describe the tools acquired for market analysis, and to illustrate their capabilities and complementarities with an example of analysis.

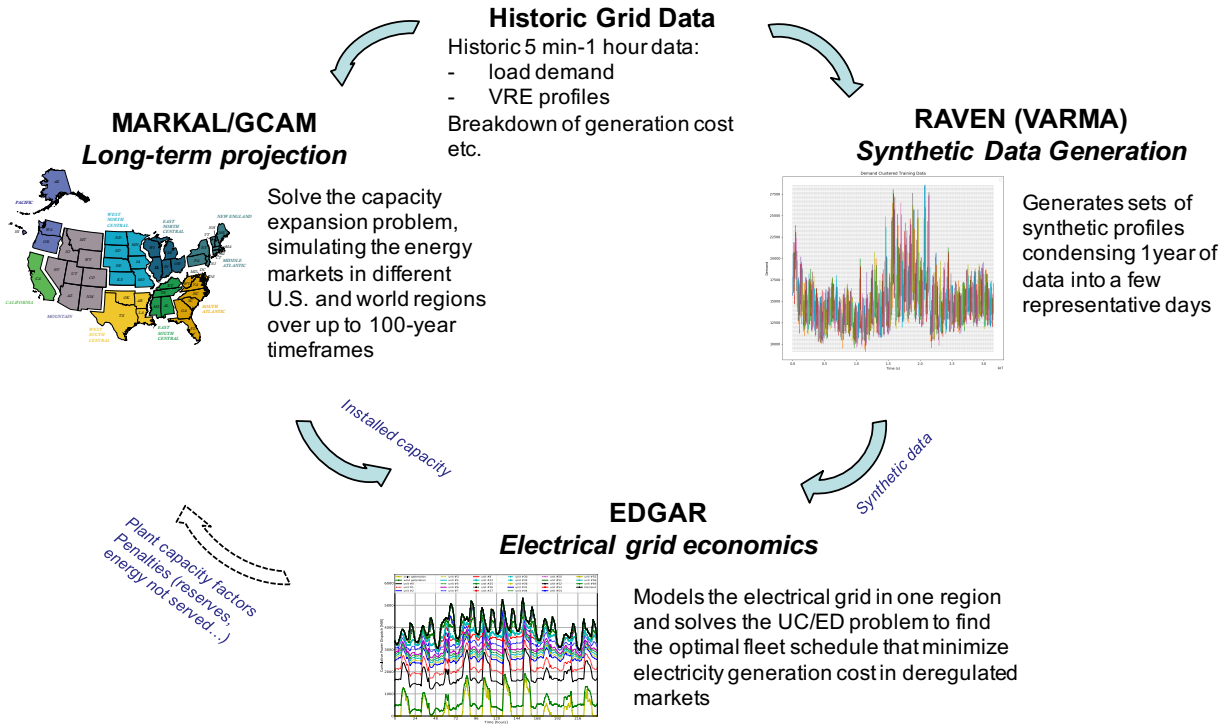

Figure 1. Methods developed and codes acquired for electricity market analysis.

The GCAM and MARKAL models solve the capacity expansion problem, simulating the energy markets in different U.S. and world regions over more than 100 years. These energy systems *models* calculate the least cost set of technologies over time that satisfy the specified demands within the bounds of the user-defined constraints (technological, political, etc.). While these codes provide a scenario-based long-term perspective, they rely on simplified assumptions (based on 1 to 5-year time steps) to account for the daily variations in electricity prices and unit generation, those need to be verified using smaller time-frames market analysis code systems.

The EDGAR (Economic Dispatch Genetic AlgoRithms) code is being developed under the SA&I Campaign to solve the combined Unit Commitment and Economic Dispatch problems to find the optimal schedule of a fleet of generating units to meet the forecasted grid demand over the next day in deregulated markets with an hourly time resolution. It allows modeling a specific competitive grid market and to assess the impact of policies and technologies on the revenue of an individual unit. Alternatively, it can be used

to demonstrate the feasibility of technology deployment scenarios proposed by capacity expansion codes. EDGAR is being developed within the Campaign since it has the unique capability of accurately modeling nuclear units by accounting for nonlinear dynamics to model xenon poisoning, length of hot and cold start-up sequences, etc. The capabilities of EDGAR were significantly expanded in FY 2018 & 2019, by improving its code structure and computation performance, adding physics modeling for xenon reactivity effect in nuclear reactors, optimizing the renewable curtailment, and enabling deterministic assessment of the reserve requirement.

EDGAR relies on sets of load demand, wind and solar generation data with a one-hour time-step. Those can be generated out of historic data using the VARMA (Vector Auto-Regressive Moving Average) model in RAVEN (Risk Analysis Virtual ENvironment), to provide a statistical understanding of the expected range of performance of a market system. The VARMA algorithm generates year-long hourly-resolution synthetic data histories. Further, RAVEN can collapse the synthetic histories to reduced-size truncated histories that are statistically representative of the full year modelled, while maintaining the correlations within different sets of data. It provides some unique capabilities that were developed within the Campaign to perform statistical sampling on both capacity expansion and unit-commitment/economic dispatch analyses to determine the best configuration for any likely weather scenario. To deliver this, segment clustering was implemented in the VARMA algorithm in FY 2019, together with two variance handling methods (segmentation and distribution preservation) conceived to better capture the distribution values from the training data.

This full suite of market economic analysis codes was used to model the New York ISO region in order to demonstrate the capabilities acquired and build expertise within the Campaign in daily market analysis. This exercise was especially useful to help better understand what are the specificities of the different market modeling codes acquired and developed, what are the assumptions these codes rely on, and where are the remaining gaps in our tools those need to be addressed. NY-ISO was selected because it is a mostly deregulated market with significant fraction of nuclear and wind generations foreseen in 2050. Long-term scenarios from capacity expansion codes can be used to drive simulations with daily market analysis codes following a few preliminary steps. First, consistent grid plant data and cost data must be gathered. Second, historic load demand and renewable generation data with fine time resolution should be obtained and processed through VARMA to condensate the full year statistical information into a few representative days. Daily market analyses are then performed with EDGAR to schedule a representative fleet of units (40 modeled for NY-ISO) on the reference time-point (2015) to demonstrate convergence of the results obtained, and their sensitivity to different sets of synthetic data generated with VARMA. Finally, similar analysis can be performed on the long-term (2050) scenario produced by GCAM to assess the feasibility of the deployment scenario provided. This procedure allows confirming the installed capacity is sufficient to meet the demand, that reserve requirements are met at every time of the year, and that the VRE generation would not lead to excess generation if their curtailment is allowed. The curtailment rate can then be provided back to the capacity expansion code in order to improve its model.

Consequently, the daily market analysis codes acquired by the SA&I Campaign enable analyses that are complimentary to the global and regional energy market studies performed with GCAM and MARKAL. In particular, some of these codes are developed within the campaign as they provide unique capabilities for accurately modeling nuclear units and accounting for uncertainties in load and renewable generation data. Current applications discussed in this report are limited to testing and developing analysis experience while additional work is underway to improve accuracy, performance, and to keep extending the type of analyses enabled with daily market modeling codes. Future effort will also focus on applying this approach to additional U.S. regions.

CONTENTS

FIGURES

TABLES

ACRONYMS

ED: Economic Dispatch

EDGAR: Economic Dispatch Genetic AlgoRithm

GCAM: Global Change Assessment Model

MILP: Mixed-Integer Linear Programming

NY-ISO: New York Independent System Operator

RAVEN: Risk Analysis Virtual ENvironment

UC: Unit Commitment

VARMA: Vector Auto-Regressive Moving Average

VRE: Variable Renewable Energy

SYSTEM ANALYSIS AND INTEGRATION CAMPAIGN DAILY MARKET ANALYSIS CAPABILITY AND RESULTS

1. INTRODUCTION

Nuclear energy plays an important role in the U.S. energy mix that will likely need to be maintained or strengthened to achieve significant greenhouse gas reduction. However, maintaining the nuclear portfolio becomes increasingly challenging in the current U.S. energy market since the low price of natural gas and the penetration of subsidized and low-marginal cost variable renewable electricity (VRE) are affecting the profitability of nuclear units. In this context, building new nuclear power plants will require increased competitiveness with reduced capital and O&M costs and increased revenues enabled by changes in market policies or increased flexible operation. Within the U.S. Department of Energy, Office of Nuclear Energy, the System Analysis and Integration Campaign has been acquiring the capability to model energy market economics in order to assist decision makers and nuclear utilities. The methods developed and codes acquired are displayed in Figure 1-1.

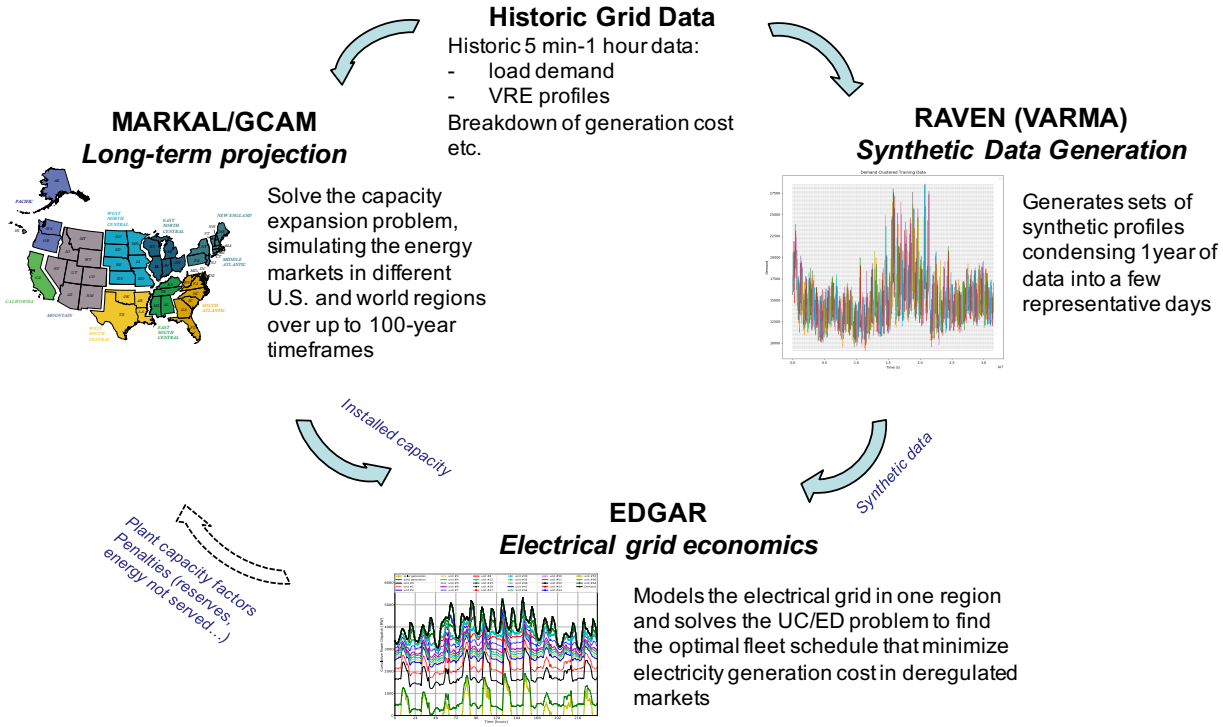

Figure 1-1. Methods developed and codes acquired for electricity market analysis.

The GCAM and MARKAL models solve the capacity expansion problem, simulating the energy markets in different U.S. and world regions over more than 100 years. These energy systems *models* calculate the least cost set of technologies over time that satisfy the specified demands within the bounds of the user-defined constraints (technological, political, etc.). While these codes provide a scenario-based long-term perspective, they rely on simplified assumptions (based on 1 to 5-year time steps) to account for the daily variations in electricity prices and unit generation. This is why shorter timeframe economic market modeling codes were acquired to complement these analyses.

The EDGAR (Economic Dispatch Genetic AlgoRithms) code solves the combined Unit Commitment and Economic Dispatch problems to find the optimal schedule of a fleet of generating units to meet the

forecasted grid demand over the next day in deregulated markets. It allows modeling a specific competitive grid market and to assess the impact of policies and technologies on the revenue of an individual unit. Alternatively, it can be used to demonstrate the feasibility of technology deployment scenarios proposed by capacity expansion codes.

EDGAR relies on sets of load demand, wind and solar generation data with a one-hour time-step. Those can be generated out of historic data using the VARMA (Vector Auto-Regressive Moving Average) model in RAVEN (Risk Analysis Virtual ENvironment), to provide a statistical understanding of the expected range of performance of a market system. The VARMA algorithm generates synthetic data histories to build truncated histories that are statistically representative of the full year modelled, while maintaining the correlations within different sets of data.

The objective of this report is to describe the capabilities acquired and developed within the SA&I Campaign together with the missing capabilities required to analyze economic scenarios of interest to DOE-NE. The different codes are described in Section 2 together with the developments performed in FY 2018 & FY 2019 under the SA&I Campaign. These different codes allow analyzing the energy markets economics in different types of areas (from a state, a utility, a country, the world) and time-frames (from minutes to more than 100 years) to complement each other and work towards a common analysis. In particular, Section 3 demonstrates with some example applications how the regional market analysis codes are used to drive simulations scenarios with daily market codes. For demonstration purposes, the economic analysis of the New York State is performed to assess the impact of different policies on nuclear competitiveness in the 2050 timeframe.

2. Codes and Methods Description

The objective of this section is to describe the capabilities acquired and developed within the SA&I Campaign together with the plan for future developments.

2.1 Global and Regional Market Codes

The SA&I Campaign uses tools to model energy supply and demand on different time scales. In the longest timeframe, market models like GCAM and MARKAL provide insights on general market direction under different socio-economic and policy scenarios. Timeframe for model simulations are typically multi-decadal in 5-year time steps and simulations into the 22nd century are necessary because of the long technical life of energy systems. These models do not have predictive value, but are useful in the exploration of alternative future socio-techno-economic pathways. In practice, they can inform on potential future deployments and retirements of the different types of generation facilities. The models are heavily data-driven, and use different types of data, such as:

- Macroeconomic data: projections of population (by state/region) and gross domestic product. Macroeconomic data are compiled by government agencies such as the Census Bureau, the Social Security Administration or the Congressional Budget Office.
- Projections of energy demand: the demands (either for energy or for energy services) can be determined endogenously or exogenously based on macroeconomic data.
- Technology data: current and potential future "technologies" have to be specified in detail. Examples of inputs needed to describe each technology include mass flows, energy expenditure and/or production, emissions generated, capacity factors, useful life, … Examples of technologies are a uranium mine ("resource" technology), an LWR ("conversion" technology) or a gasoline passenger car ("demand" technology).
- Costs: the costs of raw materials, their conversion and/or transportation, as well as capital and O&M costs of all technologies. On the nuclear side, the SA&I Campaign periodically updates the Advanced Fuel Cycle Cost Basis report, which includes costs of materials and technologies involved in the nuclear fuel cycle [1]. For non-nuclear technologies, the successive Annual Energy Outlook (AEO) reports from the Energy Information Administration are heavily used. The AEO 2019 [2] provides modeled projections of domestic energy markets through 2050. Beyond 2050, data comes from modeler judgement and/or expert elicitation when possible; understandably, the uncertainty in the input data increases for dates further in the future.
- Constraints: constraints can be technological (e.g. a limit in the growth rate of deployment of a certain technology), or can be used to model different policy assumptions (e.g. a carbon tax).

From this input information, the market models project demand and assess options for supply to meet the demand over every time period modeled. To account for the significant uncertainty associated with long-term projections of model inputs, analyses are conducted based on multiple scenarios that consider variations in demand, fuel costs, policies, etc. Examples of policy implementations include the exploration of advanced technologies, technology portfolio standards, subsides or taxes on specific technologies, and carbon emissions mitigation.

2.1.1 GCAM

GCAM is a global integrated model [3] that represents the behavior of, and complex interactions between, five systems: energy, water, agriculture and land use, economy, and climate. GCAM has been under development for over 35 years. Throughout its lifetime, GCAM has evolved to address an expanding set of science and assessment questions. The original question to which the model was applied was the magnitude of mid-21st-century global emissions of fossil fuel CO_2. Over time, GCAM has expanded its scope to include a wider set of energy producing, transforming, and using technologies; emissions of non-CO_2

greenhouse and air pollutant gases; agriculture and land use; water supplies and demands; and physical Earth systems. It is increasingly being used in multi-model, multi-scale analysis, in which it is coupled to other models with different foci and often-greater resolution in key sectors. GCAM has been used to produce scenarios for national and international assessments ranging from the very first IPCC scenarios through the present Shared Socioeconomic Pathways. Hundreds of papers have been published in peer-reviewed journals using GCAM and the model continues to be an important tool for scientific inquiry. GCAM is also a community model being used by researchers across the globe, creating a shared global research enterprise.

For the analyses performed for the SA&I Campaign, the economic and energy system representation of the United States in the GCAM model is disaggregated from a single national representation to one that includes all 50 states. The electric power sector, in particular, is fully represented for each state to assess the contribution of alternative energy technologies for electricity generation at the state level. State-level energy projections provide a more detailed and nuanced understanding of long-term national energy needs. In particular, state-level nuclear energy use at present and into the future is simulated. This capability highlights potential regional issues of energy technology competition of alternative nuclear energy technologies within the context of the broader national and global energy system.

2.1.2 MARKAL

A MARKAL model [4] provides a framework, based on the Reference Energy System concept (RES, depicted in Figure 2-1), to connect existing and potential energy carriers and conversion technologies from initial resource extraction to ultimate consumption by consumers. The model solution identifies the lowest cost combination of energy resources and technologies that meets energy service demands over the entire modeling time period, subject to specified constraints. Environmental emissions, resource use, capital investments and operating costs of energy technologies are all tracked. In other words, the model determines the market share of each technology, which depends not only on its individual characteristics (technical, economic, and environmental), but also on the availability and cost of the fuels (from the supply side) it uses.

BNL maintains two separate models of the U.S. energy system, a single-region model and a multi-region (10-region) model, shown in Figure 2-2.

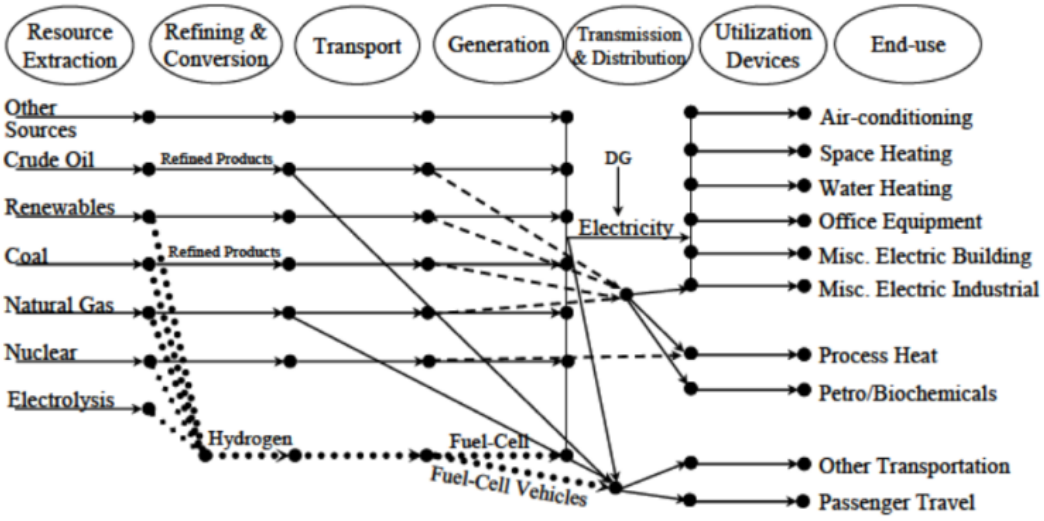

Figure 2-1. Reference Energy System Example.

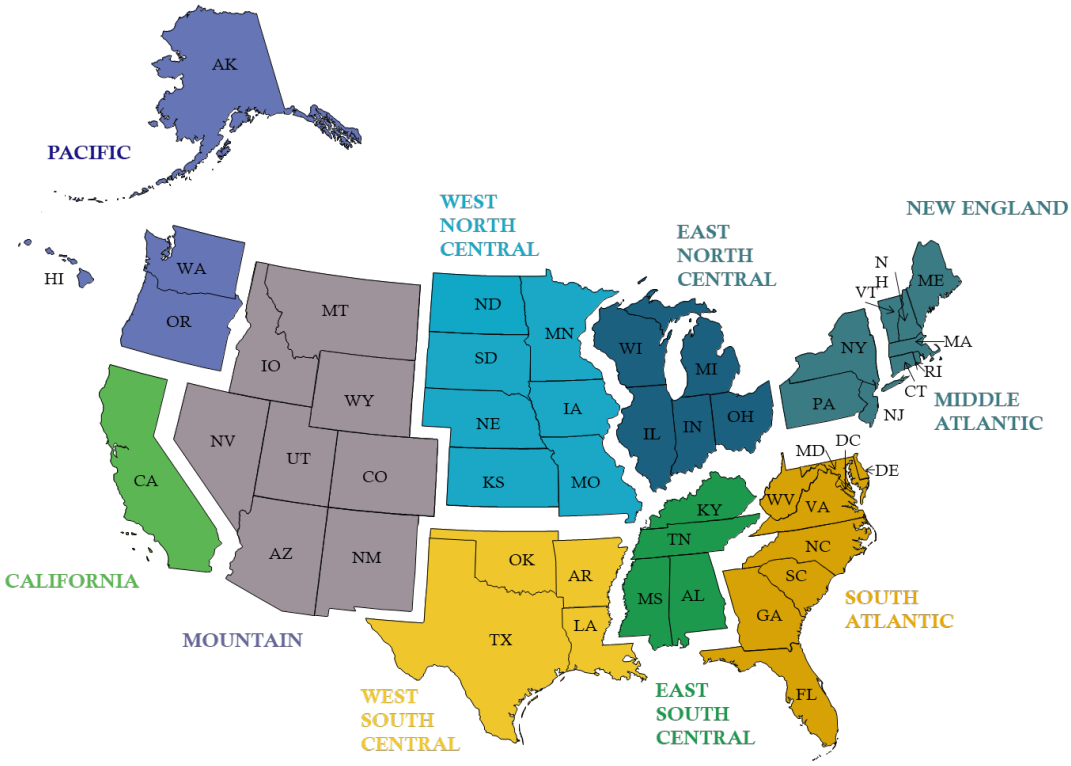

Figure 2-2. MARKAL 10-region Model.

2.1.3 GCAM vs. MARKAL

While both GCAM and MARKAL are used by the SA&I Campaign to inform on the capacity expansion problem, there are important differences between the models. A critical difference is in the kinds of questions that these models are intended to explore, which leads to differences in the modeling framework and approaches for conducting analysis. The objectives and intended applications of the GCAM is broader than that of MARKAL. GCAM is referred to as an integrated assessment model, which incorporates agriculture, land-use, water, and climate modeling capabilities in addition to the representation of the energy and economic systems. MARKAL and the family of MARKAL models are more specifically focused on the energy system and changes to the energy system dependent on economic growth projections.

While GCAM and MARKAL use similar types of data, the complexity and abundance of data used in the models, as well as different modeling assumptions/priorities, leads to differences in their input datasets. An example of different modeling approach is the treatment of VRE sources. MARKAL uses the commonly-used time slice approach, where the availability of VREs is captured according to season (3) and time of day (4), while GCAM does not use time slices for representing electricity load duration curve. Instead, the additional cost of variable energy integration is estimated by calculating the backup capacity requirement and associated cost. For wind and solar without energy storage, a backup capacity requirement is calculated as a function of the share of wind and solar energy of total electricity generation.

Underlying differences in the structure of the models also affect model results. In particular, GCAM and MARKAL have different approaches for determining the technology choice behavior. GCAM utilizes the discrete choice method, a probabilistic approach for determining the choice behavior among a set of discrete alternatives. In this method, a choice function, such as the logit function, is utilized along with the choice indicator, such as cost or profit rate, to determine the market share of available technology options. In this

approach, the single best choice (determined by the choice indicator) captures a large share of the market, but does not necessarily capture the entire market. Alternative higher cost or less profitable technology options may capture portions of the market reflecting consumer preferences, local variations in costs, and other non-observable factors. Technology options typically share the market unless the choice indicator is significantly divergent, in which case the best choice may capture the entire market. Policies for promoting or hindering specific technologies are achieved through the adjustment of the choice indicator, such as by adding a subsidy or tax to the technology cost.

MARKAL utilizes a linear-programming approach to the technology choice behavior. In this approach, the technology choice is determined by an optimization problem as specified by the objective function, decision variables and a set of constraints. The objective function is typically a cost minimization or a profit maximization formulation, and the decision variables represent the choices made by the model, such as the level of technology deployment. The set of constraints applied to the linear relationships constrain the optimization problem and play an important role determining the outcome of technology shares. The least cost or the most profitable technology typically dominates market share, which is mitigated by the set of technology constraints. Constraints serve multiple functions, such as for setting deployment levels for technology policies or for capturing other non-observable factors that influence the technology choice.

2.2 Daily Market Codes

Daily market codes were acquired or further developed by the SA&I Campaign in FY 2018 to complement the Global and Regional Market analyses performed with MARKAL and GCAM. They perform economic market modeling on shorter timeframe providing essential information to account for some changes in daily load demand profiles or VRE generation profiles. This section describes the methods in EDGAR and RAVEN/VARMA together with the work accomplished in FY 2018 & 2019 by the SA&I Campaign on these codes. Since EDGAR and RAVEN/VARMA are new acquisitions from the Campaign and significant development work was performed to enable analysis requirements, their description is detailed in this section to serve as a reference in the future.

2.2.1 EDGAR

2.2.1.1 Introduction

The EDGAR (Economic Dispatch Genetic AlgoRithm) code is an electrical grid modeling tool developed in the Nuclear Science and Engineering (NSE) division of the Argonne National Laboratory (ANL). It estimates the hourly power generated by each unit within a fleet in a deregulated market together with the electricity production costs by solving the unit commitment (UC) and economic dispatch (ED) problems following the approach described in Figure 2-3. The main novelty of EDGAR is its capability of accounting for highly non-linear constraints such as xenon-related constraints on nuclear reactors operation, while those are challenging to model using traditional algorithms based on Mixed-Integer Linear Programming (MILP) schemes.

The UC problem in electrical power systems is a large family of mathematical optimization problems where the electricity generation of a set of generating units is coordinated and optimized to meet the energy demand at minimum energy production cost in a deregulated market. The continuous balancing of supply and demand in the power grid requires that each variation in the consumption needs to be instantaneously matched by a corresponding variation in the production. As a result, the solution of the UC problem determines which generating units in a given fleet need to be committed to meet the expected load over the next day. Basically, the UC problem solution provides a schedule for all the units in a fleet for one day, which specifies which units are on and off at every time-step (typically 1 hour).

The ED problem is the short-term determination of the optimal output of previously-committed generating units to meet the load demand and the minimum reserve requirements, by minimizing the whole production

cost and by meeting the corresponding transmission and operational constraints. In particular, the ED problem solution specifies the power outputs of each active unit on a time-step basis.

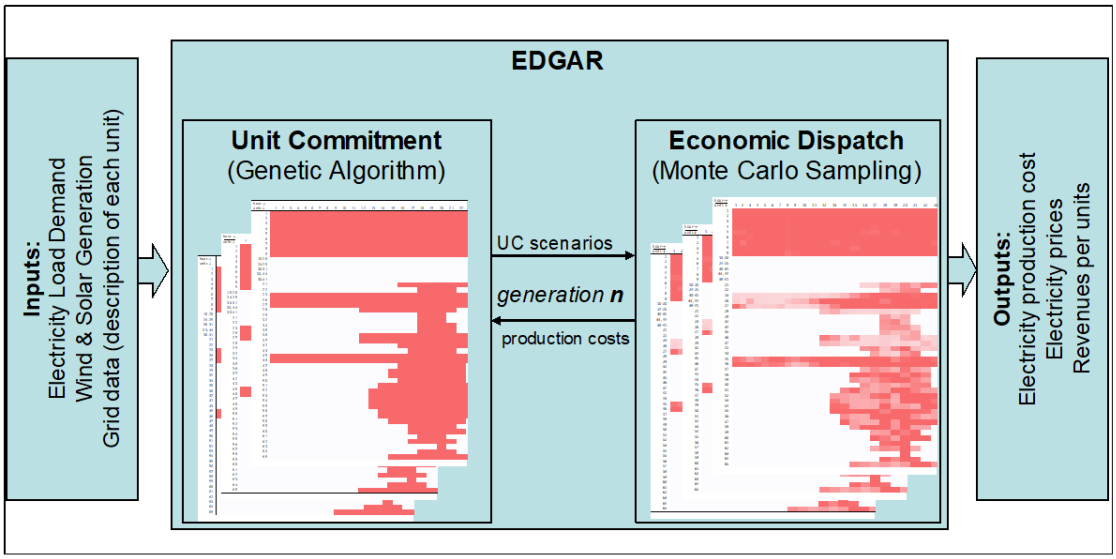

Figure 2-3. Description of the EDGAR code.

As for the adopted algorithm, the EDGAR code minimizes the power generation costs of the system for each day by using a Genetic Algorithm (GA) for the UC optimization, and solves the ED problem for each UC scenario by means of a constrained Monte Carlo sampling approach. EDGAR optimizes the grid dispatch while taking into account reserve requirements and ensuring every unit is operating while meeting all its operational constraints.

2.2.1.2 Methods overview

2.2.1.2.1 Unit Commitment

The EDGAR code relies on a GA for the UC optimization with the goal of minimizing the production cost of the fleet on the day ahead. GAs belong to the family of evolutionary algorithms, which mimic aspects of Darwin's theory of natural selection. Upon completion, a GA will return an optimal or near optimal individual after a number of iterations [5]. In literature, several applications of GAs to solve the UC/ED constrained optimization problem can be found [6, 7].

The GA implemented in EDGAR follows this generic approach as shown in Figure 2-4. The basic idea is that a set of feasible, tentative UC schedules (*chromosomes*) on a given day is generated at random in the initial step. Each chromosome represents the schedule of the whole fleet of units over the entire day, and is characterized by an optimized ED scenario, which has a daily cost associated with it. As a consequence of the random generation process, these potential fleet schedules might not be characterized by low costs. Evolution proceeds by improving the initial population by the use of genetic operators, i.e., crossover and mutation. In particular, cheaper scenarios are selected more often and recombined to generate cheaper scenarios up to converge to the global minimum. Mutations are applied to some solutions to increase the genetic diversity. The new generation of UC scenarios has then its daily costs evaluated, and the process is repeated through generations until the termination criterion is fulfilled. Once the condition is met, the refined ED calculation is performed to compute the hourly cost and price profiles, before moving on to simulating the following day.

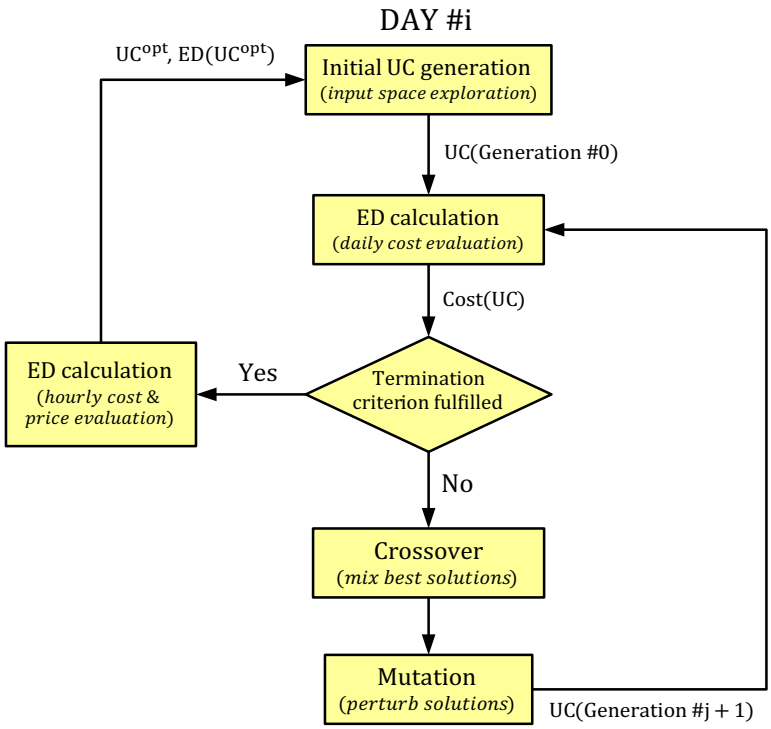

Figure 2-4. Genetic Algorithm procedure implemented for UC/ED calculation in EDGAR.

As shown in Figure 2-4, the first step consists in generating tentative UC schedules. One of the main challenges is to inherently respect the operational constraints. In particular, many generators have constraints on times between start-up and shut-down, i.e., after the unit is started-up, it needs to be committed for a certain number of hours before it shuts-down, and vice versa. Every time the tentative UC solutions are generated, the minimum up/down time constraints need to be accounted for. At the end of a simulated day, the hourly state of every unit is used to automatically update the initial conditions for the evaluation over the next day.

The key ingredient of the evolution process is the definition of a suitable metrics to establish which individuals should have a higher probability of being allowed to multiply and reproduce, and which individuals should have a higher probability of being removed from the population. The GA evaluation metrics is called *fitness* function. Fitness functions are very problem-specific. Since the goal is to minimize the production cost of the fleet on the day, the EDGAR code fitness definition is related to the overall cost associated with the corresponding UC scenario, i.e., the cheapest UC scenarios have better chances of being selected. Logically, the adopted metrics would be different if the optimization goal was different, e.g., instead of minimizing the overall operation cost, one might want to find the UC scenario which maximizes the profitability of the nuclear units.

After the available UC individuals are ranked according to the chosen metrics, pairs of them are chosen as parents based on their fitness. It is a stochastic process of selection, which draws individuals on the basis of their performance as compared to other individuals, keeping the best scenarios and maintaining diversity of solution. This principle is essential in preserving the genetic diversity and helps keep the search away from local optima [5]. The parent schedules are recombined through cross-over operations to form a new generation of UC scenarios, i.e., the genetic material of two parents is combined by swapping a part of one parent with a part of the other, and two child UC scenarios are created. If the portions of chromosomes involved in the pairing process were chopped vertically and then adjoined, the minimum up-down time constraints might be violated. This does not happen if the portions of the parent string are chopped

horizontally and then adjoined. Since the initial population is generated by respecting the minimum time constraints, the application of this version of the crossover operator ensures meeting the constraints. The adopted approach is described graphically in Figure 2-5.

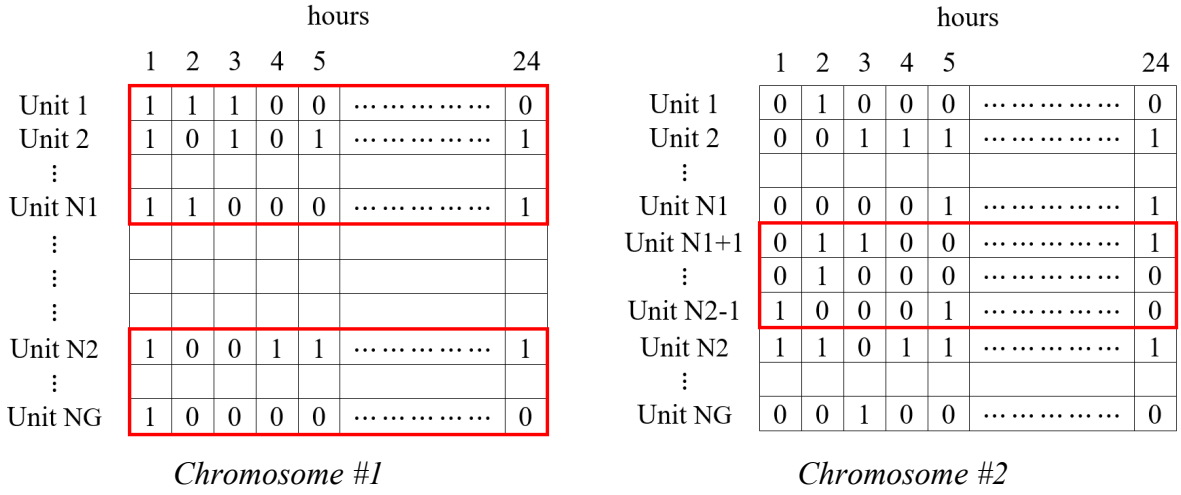

Figure 2-5. Cross-over approach.

After crossover has occurred, each one of the produced children undergoes mutation with a certain probability. With respect to the crossover operator, which exchanges genetic material between two individuals to create a fitter offspring, the mutation operator changes a small portion of an individual. In EDGAR, when a UC scenario is selected for mutation, a certain unit is randomly selected, and the corresponding schedule is altered. In particular, it can undergo through four types of possible mutations:

- a committed unit has its daily profile re-sampled
- a previously committed unit is shut down
- a new unit is committed (with committed profile sampled)
- a committed unit is replaced by a non-committed unit (with committed profile sampled)

Finally, the ED calculation is performed on every child (new UC scenario obtained in the new generation) following the method described in the following section. The best scenarios are then selected for generating a new generation of children. Several generations are then evaluated until a convergence or termination criterion is fulfilled. For instance, this criterion can be based on the number of generations. Alternatively, it can be imposed that if the fitness of the best individual has not experienced any improvement within the last 5-10 generations, the GA run can be interrupted. Upon termination, the best UC scenario is then selected for refined cost evaluation (hour-by-hour cost and price) before starting next day's evaluation. The "clearing price" is computed with EDGAR following the UC calculation, and is provided as an output of the code. This is the highest bid accepted during the day-ahead electricity auction, which sets the price that is going to be paid to all the committed generators whose bids are accepted. Quantitatively, the clearing price is equal to the marginal cost of the most expensive unit that was selected to be committed.

2.2.1.2.2 Economic Dispatch

Each UC scenario is characterized by a power generation cost which ultimately determines its fitness. A selection process combines UC scenarios to generate a new generation of tentatively cheaper UC scenarios. The overall power generation costs and the power dispatched by each generating unit are derived as solutions of the ED problem by adopting a constrained Monte Carlo sampling approach. The imposed constraints comprise output power boundaries for the generating units, minimum up and down times, limits

on the speed and the frequency of the operational transients, supply/demand balance over the simulated time window as well as provisions for different types of operating reserves. The solution of the ED problem is obtained by minimizing the total cost of meeting the demand for the next day, as shown in Eq. (1). The current model has an hourly time resolution, and the optimal solution is obtained by seeking the cheapest solution over a 24-hour time horizon.

$$min\{total\ cost\} = min\left\{\sum_{i,t} cpg(i,t) + \sum_{i,t}[C_{SU}(i,t) + C_{SD}(i,t)] + \sum_{t}[C_{ENS}(t) + C_{EOP}(t) + C_{RNS}(t) + C_{NRNS}(t) + C_{RGNS}(t)]\right\} \quad (1)$$

The first term represents the summation of the single unit power generation costs ($cpg(i,t)$), which are defined as:

$$cpg(i,t) = NLC(i) \cdot u(i,t) + MC(i) \cdot P(i,t) \quad (2)$$

where $NLC(i)$ is the no-load cost for i-th unit [$], $MC(i)$ is the marginal cost of the i-th unit [$/MW], and $P(i,t)$ is the electrical power output of i-th unit at time t [MW]. The second term represents the summation of the unit start-up ($C_{SU}(i,t)$) and shut-down costs ($C_{SD}(i,t)$), [$].The last term represents the penalties associated with unmet demand and reserves. In particular, with respect to time t, $C_{ENS}(t)$ represents the cost of unserved energy [$], $C_{EOP}(t)$ represents the cost of over-produced energy [$], $C_{RNS}(t)$ represents the cost of spinning reserve not served [$], $C_{NRNS}(t)$ represents the cost of non-spinning reserve not served [$], and $C_{RGNS}(t)$ represents the cost of Regulation not served [$]. It should be noted that the fixed costs (e.g., capital cost, most of the fixed O&M costs, etc.) do not impact the solution that seeks the most cost-effective way to schedule a group of power-producing units to meet demand and reserves profiles over the next day.

As for the ED problem optimization, for each UC scenario, a large number of power histories (ED scenarios) are sampled for each one of the committed generating units. For each ED scenario, certain committed units are randomly sampled to increase their electrical power output (i^{up}) and other ones to reduce it (i^{down}) within the authorized limit in terms of output and ramp rates. The power sampling is calculated using the following equations:

$$P(i^{up}, t+1) = P(i^{up}, t) + a \cdot \Delta P_{max}(i^{up}, t) \quad (3)$$

$$P(i^{down}, t+1) = P(i^{down}, t) + a \cdot \Delta P_{min}(i^{down}, t) \quad (4)$$

where a is a random number sampled from the distribution $U[0;1]$. A major outcome of this approach is to ensure that the constraints on the maximum power variation allowed are inherently met, and the operational constraints which might be induced by the unit power dispatched are accounted for. From this standpoint, let's consider the case of nuclear units operated in a load-following mode. When the power level of a nuclear power plant is reduced, there are physics-induced constraints that potentially prevent the unit from ramping up to nominal power conditions, as further discussed in Ref. [8]. Once the power level is reduced, a thermal reactor unit may need to be operated at lower power for a number of hours due to the xenon poisoning level, which depends on the amplitude of the imposed power drop. The implementation of this physical effect is particularly troublesome in MILP optimization algorithm [9]. On the other hand, the implementation of this non-linear constraint in the proposed Monte Carlo sampling algorithm is pretty straightforward as explained in Section 2.2.1.3.2.

At this point, the contributions of the different units to the reserve need to be evaluated for each UC scenario. EDGAR receives the hourly reserve requirement as an input of the code. However, for analyzing new grid setups, one needs to be able to estimate these reserve requirements as discussed in Section 2.2.1.3.4. Assuming one knows the hourly reserve requirements, six different kinds of reserves are considered by EDGAR: spinning ($sr(i,t)$), regulation-up ($rgu(i,t)$), regulation-down ($rgd(i,t)$), non-

spinning when the unit is committed ($nsrn(i,t)$), down ($dr(i,t)$), and non-spinning when the unit is not committed ($nsrf(i,t)$). The contributions provided by each unit are governed by the following constraints:

$$\begin{aligned} sr(i,t) &\leq [sr_\tau \cdot MSR(i)] \cdot u(i,t) \\ rgu(i,t) &\leq [reg_\tau \cdot MSR(i)] \cdot u(i,t) \\ rgd(i,t) &\leq [reg_\tau \cdot MSR(i)] \cdot u(i,t) \\ nsrn(i,t) &\leq [nsrn_\tau \cdot MSR(i)] \cdot u(i,t) \\ dr(i,t) &\leq [sr_\tau \cdot MSR(i)] \cdot u(i,t) \\ nsrf(i,t) &\leq [QSC(i)] \cdot [1 - u(i,t)] \end{aligned} \tag{5}$$

where sr_τ, reg_τ, $nsrn_\tau$ are the ancillary service characteristic time constants [min], $MSR(i)$ is the maximum sustained ramp for i-th unit [MWh/min], and $QSC(i)$ is the quick start capacity for i-th unit [MWh]. To meet these constraints, the contribution of every unit is obtained by multiplying the corresponding upped bounds by random numbers sampled from the distribution $U[0;1]$. Finally, the sampled contributions are weighted by the corresponding unit commitment schedule, i.e., $u(i,t)$.

Once the power outputs from each one of the committed units are sampled, and as are the corresponding reserve contributions, the instantaneous balance with the prescribed values are verified by means of an internal loop. Thanks to this two-step procedure within the ED, the obtained power histories is assured to meet the global grid demand and to respect the singular unit operational constraints. The final ED scenario is verified to be feasible and it is characterized by a certain cost, which is evaluated according to the above-defined cost function (Eq. (1)). At this point, once all the sampled scenarios are collected, the cheapest one is sorted out, and the fitness representative of the corresponding UC configuration is derived.

2.2.1.3 EDGAR developments in FY18 & FY19

Major developments were accomplished in FY 2018 and early FY 2019 on the EDGAR code system to improve its usefulness to the SA&I Campaign.

2.2.1.3.1 Structural code improvements

The EDGAR code was translated from MATLAB into Python to facilitate the portability to different machines and platforms (Unix Clusters, MacOSX and Windows Desktops). The development of the code is performed within the Gitlab repository (hosted by MCS division at Argonne), which allows efficient collaborative development and implementation of a quality assurance strategy. A parallel mode version of EDGAR was implemented, which allows significantly enhancing the computational performance of the code, which was found to be critical for solving a full UC/ED problem with sufficient convergence.

Significant effort allowed re-organizing part of the EDGAR code to move away from calculation script to a more formal code structure (with a proper input, an executable, …). This was accomplished by leveraging the "Workbench" universal code interface (Open Source) developed by the NEAMS program to apply it to EDGAR, which provides the following benefits:

- Optional access to the Workbench user interface with input validation and auto-completion features (plus some result visualization that we may set up in the future) – the user still has the option to run EDGAR as a standalone code outside of the Workbench user interface.
- Input documentation that is automatically generated from the input structure description.
- Structures for unit tests, those were implemented for better quality assurance strategy and are for checking the pre-, post-processing and execution logics. In particular, regression tests are

implemented to verify the results of the economic dispatch and of the unit commitment procedures with reference solutions (obtained with the AMPL code on the APS grid [10]).

In addition, the Workbench [11] provides a common user interface for model creation allowing for its integrated codes to communicate and work together with limited coupling development [12]. In particular, the coupling between Dakota [13] and EDGAR was used in FY 2019 for demonstration purposes to enable solving "green-field" capacity expansion problem to optimize the number and capacity of each generating unit deployed. The Dakota software, which is already integrated within the NEAMS Workbench, is a sensitivity analysis/uncertainty quantification (SA/UQ) and optimization toolkit maintained by Sandia National Laboratory. This type of problem can also be performed by coupling EDGAR to RAVEN as further discussed in Section 2.3.3.

2.2.1.3.2 Nuclear load following modeling

To evaluate the impact of the load-following operation on the nuclear unit profitability, the physics-induced limitations affecting unit flexible operation capabilities need to be modeled. From this standpoint, a dedicated constraint describing the xenon poisoning-induced effects has been implemented in EDGAR as detailed in [14].

The xenon constraint is considered by EDGAR for each individual nuclear unit with a reactivity balance criterion applied to assess the feasibility of the sampled power evolution at every time step. As shown in Figure 2-6, at day j, the nuclear unit ramp-up after a power drop (ΔP) can be performed only if the available reactivity margin ($\Delta\rho_{marg}(j)$) is higher than the xenon negative reactivity insertion ($\Delta\rho_{Xe}(\Delta P)$). Otherwise, it will be necessary to wait for the xenon to decay before starting a ramp to higher power level, which can typically require a few hours since the half-life of ^{135}Xe is 9.2 hours.

Finally, an additional constraint to be considered for the nuclear units is the one imposing the nuclear unit to stay at reduced power level for two hours after a power drop [15] to represent the need of limiting the thermal stresses induced on the internals.

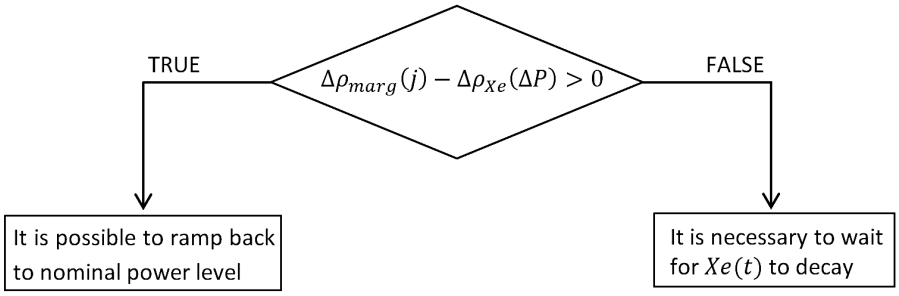

Figure 2-6. Implemented reactivity balance criterion which governs the reactor operation [9].

2.2.1.3.3 VRE curtailment

The possibility of curtailing renewable energy source contribution was implemented in EDGAR in FY 2019. The curtailment is the reduction in the output of a generator from what it could otherwise produce given available resources, typically on an involuntary basis [16]. Because wind and solar generators have substantial capital costs but no fuel or flexible costs, the grid operators try to exploit their contributions as much as possible. Operator-induced curtailment of wind and solar energy is becoming more diffused as wind and solar energy penetration increase, and it typically occurs because of transmission congestion or

lack of transmission access, but it can occur for a variety of other reasons, such as excess generation during low load periods, voltage, or interconnection issues.

Wind generation can be easily curtailed since turbines are typically grouped in the same location. On the other hand, solar generation ranges from large, centralized utility-scale photovoltaic power stations to distributed, residential, and commercial building integrated installations. The former contribution can generally be curtailed, the latter may not. In EDGAR, the sampled value for the wind power ($wg(t)$) ranges from zero to the maximum power availability ($W(t)$), as shown in Eq. (6). The sampled value for the solar power ($sg(t)$) is upper bounded by the maximum power availability ($S(t)$), and lower bounded by the fraction of the distributed solar power contribution (r), which is not curtailable (Eq. (7)).

$$wg(t) = U[0,1] \cdot W(t) \tag{6}$$

$$sg(t) = r \cdot S(t) + U[0,1] \cdot (1 - r) \cdot S(t) \tag{7}$$

As a consequence of the implementation of the renewable energy source curtailment capability, solar and wind power contributions were turned into unknown variables to be optimized. Accordingly, the sampled values of the non base-load unit power outputs ($P(i^{NBL}, t)$), the solar and the wind power contributions ($sg(t)$, $wg(t)$) are properly normalized to meet the portion of the load demand not covered by the must-run based load units ($D(t) - \sum P(i^{BL})$). The corresponding normalization factor is expressed in Eq. (8).

$$\frac{D(t) - \sum P(i^{BL})}{\sum P(i^{NBL}, t) + wg(t) + sg(t)} \tag{8}$$

The above-described approach was generalized in EDGAR to account for any type of fixed energy generation, simply described by their hourly generation profile and a maximum curtailment fraction.

2.2.1.3.4 Operating reserve profile generation

The objective of this work is to develop capability to generate realistic reserve profiles needed for EDGAR simulation, when historical reserve profiles are unavailable. Traditionally, power system uncertainty arises from load fluctuation and system contingencies requiring additional generation capacity to be reserved [17]. This extra generation capacity might be immediately available by increasing the power output of generators that are already connected to the power system (*spinning reserve*) or by those not currently connected to the system but whom can be brought online after a short delay (*non-spinning reserve*).

The operating reserve requirement can be estimated from load demand and some renewable profiles, since they are meant to account for the uncertainty from renewable energy source contributions, load fluctuations and outages. Generally, the methods employed by the system operators to define operating reserve requirements are deterministic. However, it may happen that complex risky situations are not covered since deterministic approaches do not in fact measure the risk. Consequently, an approach based on deterministic criteria may lead either to higher operational cost, or to excessive risk [18]. That is why system operators such as ERCOT are starting to abandon deterministic rules in favor of probabilistic methods for defining their monthly non-spinning reserve requirements. The approach consists in setting a non-spinning reserve corresponding to percentile 95 of the historical total forecast error [18].

For the sake of simplicity, in this preliminary work, a deterministic approach was adopted. In particular, the operating reserve requirements can be generated as a fraction of the largest contingency in the fleet, i.e., sudden and unforeseen shutdown of the committed unit with the largest capacity ($LC(t)$), of the daily peak load ($PL(day)$), and of the renewable energy source forecast ($RES(t)$) [19]. To assess the validity of this approach, a comparison against some available data (based on South-Western U.S. region) was performed.

The parameters reported in Eqs. (10) were adopted, and the corresponding outcomes are shown in Figure 2-7.

$$Op.Reserve(t) = a \cdot LC(t) + b \cdot PL(day) + c \cdot RES(t) \quad (9)$$

$$a = 0.35, b = 0.02, c = 0.15 \quad (10)$$

Finally, the correlation between the operating reserve requirements and the other ancillary services (i.e., regulation up, regulation down, downward reserve) were derived using available data. The total reserve requirement is equally split between the spinning reserve (SR) and the non-spinning reserve when the unit is on (NR) requirements. The other ancillary services, i.e., regulation up (RGU), regulation down (RGD), down reserve (DR), were determined as function of the spinning reserve profile. The contribution of the non-spinning reserve when the unit is off was not considered.

$$SR(t) = Op.Reserve(t) * 0.5 \quad (11)$$

$$NR(t) = Op.Reserve(t) * 0.5 \quad (12)$$

$$RGU(t) = SR(t) * 0.1 \quad (13)$$

$$RGU(t) = SR(t) * 0.1 \quad (14)$$

$$DR(t) = SR(t) * 0.1 \quad (15)$$

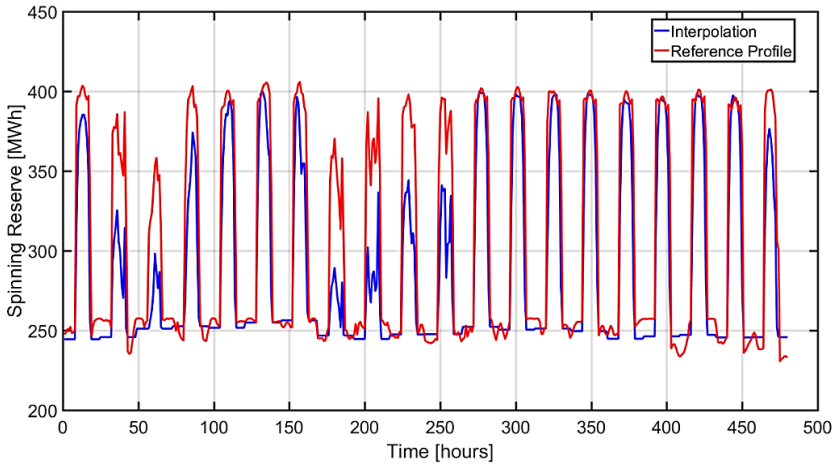

Figure 2-7. Comparison between the reference spinning reserve profile and the one retrieved by adopting the proposed deterministic correlation.

2.2.2 RAVEN

2.2.2.1 Introduction

The modeling of energy grid systems has previously been constrained to either consider long-period averages with poor time resolution, or higher time resolution but exclusively based on pre-existing data from the historical record. While useful in some applications, neither of these methods provides a basis for thoroughly exploring the general effectiveness of a particular energy grid configuration. Long-term averages ignore important short-term phenomena such as high-frequency storage usage and unit power ramping to cope with volatility of demand, while historical samples are too few in number to provide a good statistical understanding of a given grid configuration (such as in the case of ensuring Loss of Load Probability constraint). Ideally, tens of thousands of high-resolution data sets for a variety of weather and grid measurements would be available to statistically comprehend the strengths and shortcomings of a configuration; in reality, rarely are more than a dozen years of recorded data available, and rarely with more

than hourly time resolution. As a result, generation of synthetic time-dependent histories has become a field of interest in energy studies.

A recent push has been made to simulate time-dependent histories (time histories) using a variety of sophisticated mathematical models. Among these is the application of Fourier detrending together with Vector Auto-Regressive Moving Average (VARMA) algorithms [20]. In this approach, Fourier analysis is used to remove seasonal trends from time histories, and the residual signals are analyzed using VARMA. The VARMA algorithm analyzes the signals of a set of coupled histories, such as solar intensity and energy demand, by treating them as coupled noisy signals with some underlying pattern described well by autoregression (AR) and moving average (MA). Once the Fourier and VARMA descriptors are well understood for a set of signals, new synthetic samples can be produced ad infinitum. The Fourier plus VARMA algorithm has previously been implemented in the Risk Analysis Virtual ENvironment (RAVEN) [21].

The existing RAVEN algorithms (including VARMA and Fourier detrending) make it possible to produce functionally limitless numbers of identically-distributed and independent samples. This enables statistical analysis of grid configuration quality [22]. Without statistical sampling, at best a grid market analysis study (including both capacity expansion or unit-commitment/economic dispatch analyses) can determine the ideal configuration for one specific set of weather and grid behaviors from among many possible such scenarios. With statistical sampling, grid market analysis studies can determine the best configuration for any likely outcome.

2.2.2.2 VARMA description

The VARMA algorithm as contained in RAVEN consists of several algorithms that together attempt to capture and reproduce the essential characteristics of time-dependent signals such as weather and energy grid measurements. Within the RAVEN framework, the VARMA algorithm is embedded as a Reduced-Order Model (ROM), allowing it to be trained, saved, and sampled repeatedly.

The internal training algorithms for the VARMA ROM consist of two steps: seasonal detrending through Fourier analysis, and coupled residual characterization via the Vector Auto-Regressive Moving Average (VARMA) algorithm.

The Fourier detrending process involves determining the impact of Fourier signals with user-provided periods on the training history. The Fourier components $F(t)$ of the training signal are calculated as

$$F(t) = \sum_{i=0}^{k} c_i \sin\left(\frac{2\pi}{\lambda_i} t + \eta_i\right) \tag{16}$$

where i indexes the user-provided periods, c_i is a scaling coefficient, λ_i is the provided period, and η_i is a phase adjustment. Least-squares fitting is used to determine the optimal fit of all c_i, η_i to the training data, after which the full Fourier signal is subtracted from the training data, leaving a residual noise, a signal comprised of deviations from the periodic trends,

$$R(t) = T(t) - F(t) \tag{17}$$

where $T(t)$ is the original training data and $R(t)$ is the residual noise. It is worth noting that the choice of detrending periods λ is an essential component to effective VARMA training. If insufficient seasonal and diurnal trends are removed from the signal, the residual noise $R(t)$ will exhibit an inflated variance, which will translate into abnormally noisy synthetic signals and may produce unrealistic values. If over-selection of the periods occurs, then the signal may either be overfit and demonstrate nearly identical synthetic

signals, or numerical instability may present with unphysical results. RAVEN offers some tools such as the Fast Fourier Transform as guides for choosing the most important cyclic trends in a training signal.

After performing Fourier detrending, the residual signal $R(t)$ is characterized with Auto-Regressive Moving Average (ARMA) algorithms. In the event $R(t)$ consists of several signals (such as load, solar, and wind data) that may be correlated, the Vector ARMA (VARMA) extends the ARMA.

2.2.2.2.1 Single-case: ARMA

The ARMA analysis is performed ideally on a Gaussian normally-distributed signal, so the residual signal from Fourier detrending is transformed via its cumulative distribution function $f(t)$ to be distributed in a Gaussian manner,

$$y(t) = \Phi^{-1}[f(R(t))] \tag{18}$$

where Φ is the Gaussian normal cumulative distribution function (CDF) and $f(t)$ is the (empirically-derived) cumulative distribution function of the residual signal.

The Auto-Regressive (AR) portion of the ARMA seeks coefficients that best interpret y at any given time t as a function of its preceding terms in time, or lag terms,

$$y_t^{AR} = \sum_i^P \phi_i \, y_{t-i} + \epsilon \tag{19}$$

where ϕ_i are regression coefficients, P is the number of lag coefficients to include in the sum, y_{t-1} signifies the discrete value of y at the time step previous to t, and ϵ represents a random Gaussian noise component.

Restated, the AR value of any given time step t is regressively determined by the sum of components at previous time steps, plus a random noise component.

The Moving Average (MA) is a rolling average among a finite number of lag terms, and is added to the auto-regressive model to yield the ARMA,

$$y_t = \sum_i^P \phi_i y_{t-i} + \epsilon + \sum_j^Q \theta_j \epsilon_{t-j} \tag{20}$$

where Q is the number of lag terms to include in the rolling window for the moving average, and θ is a fitting coefficient associated with the moving average lag term. The coefficients ϕ, θ are fitted to maximize a likelihood function, as described in [20] and [23].

Once the fitting for c, λ in the Fourier detrending as well as the ARMA coefficients ϕ, θ have been determined, the ARMA model is trained and can provide independent, identically-distributed samples by the following process:

1. Randomly sample a number of values from a Gaussian normal distribution equal to the trained history length, and place them one after another as a history, then apply the ARMA regressive terms to calculate each value in the new history $\tilde{y}(t)$.
2. Transform the Gaussian noise to fit the original training residual noise distribution using the CDFs of both distributions to obtain the residual noise $\tilde{R}(t)$ of the new sample:

$$\tilde{R}(t) = f^{-1}[\Phi(\tilde{y}(t))] \tag{21}$$

Add the Fourier seasonal trends to the new residual noise sample to obtain the full synthetic history $S(t)$:

$$S(t) = F(t) + \tilde{R}(t) \tag{22}$$

This process can be repeated as many times as desired for statistical sampling.

2.2.2.2.1 Extension to multiple coupled variables: VARMA

Extending the ARMA to include multiple variables while maintaining the correlations within these variables is straightforward. Because the Fourier detrending is deterministic, any coupling due to season patterns is captured by applying the Fourier detrending to both signals. Remaining coupling in the residual

noise of the signals is captured by extending the AR and MA to consider not only lag terms in their own history, but lag terms in the history of the correlated signals as well. The AR component is given by

$$\vec{Y}_t = \sum_i^P \vec{\Phi}\, y_{t-i} + \vec{\mu} \tag{23}$$

where each element in $\vec{u}$ is determined by the correlation between the time series in the moving average,

$$\vec{\mu} = \epsilon + \sum_j^Q \Theta_j \epsilon_{t-j} \tag{24}$$

The dependency of the moving average term on the vector auto-regression determines the vector auto-regression equation. Letting L be the lag operator, so that $Ly_t = y_{t-1}$, then the VARMA representation is

$$\vec{Y} \sum_i^\infty \Pi_\mathrm{i}\, \vec{Y}_{t-i} = \Theta(L)^{-1}\vec{\Phi}(L)\vec{Y}_t \tag{25}$$

with the matrices Π_i given such that

$$\vec{I}_k - \sum_k^\infty \Pi_\mathrm{i} L^i = \Theta(L)^{-1}\vec{\Phi}(L) \tag{26}$$

Sampling the VARMA follows the same pattern as sampling the ARMA, resulting in a vector of correlated signals.

2.2.2.3 VARMA developments in FY18 & FY19

While the initial implementation of the VARMA was instrumental for several analyses [22], some limitations were found as it was applied to a variety of applications. Primarily, the VARMA algorithm assumes that after Fourier detrending occurs, the residual signal is distributed uniformly throughout the signal. In reality, significant time dependence was observed for many signals of interest. For example, considering a year-long energy load signal, the variance of the residual was much larger in summer than winter, and larger in winter than spring or fall. Treating this distribution as though it were uniformly time-independent resulted in much wider variance in the spring and fall and reduced variance in the summer. This wider variance translated into occasional very low demand during already-low load seasons, and occasional very high demand in already-high demand seasons. On a shorter time-scale, load residual variance is high during the transitions between days and night, but in general is smaller during the middle of the day and night. Treating this as uniform sometimes resulted in much lower loads during the night and much higher loads during the day.

The issue with significantly lower yearly maximum or yearly minimum values can be seen in the context of grid capacity optimization. The efficient operation of baseload plants assures that by definition they are very rarely curtailed. If the synthetic histories generate too many hours with low load during the year, even if a proportional number of hours with high load are also generated, the resulting optimal mix will include less baseload and more variable sources. See for example Figure 2-8, showing the original training data discussed in Section 3.2 (NYISO, load, 2018) in solid blue in the foreground and 100 synthetic samples from the basic ARMA in the background as an orange cloud. The x-axis of each figure has the months

marked at the middle (15th day) of each month. Note that the synthetic samples tend to show significantly more variance in the spring and fall than the original data, and the lowest signal values are much lower as a result. Also note the signal shows much lower values during the summer months.

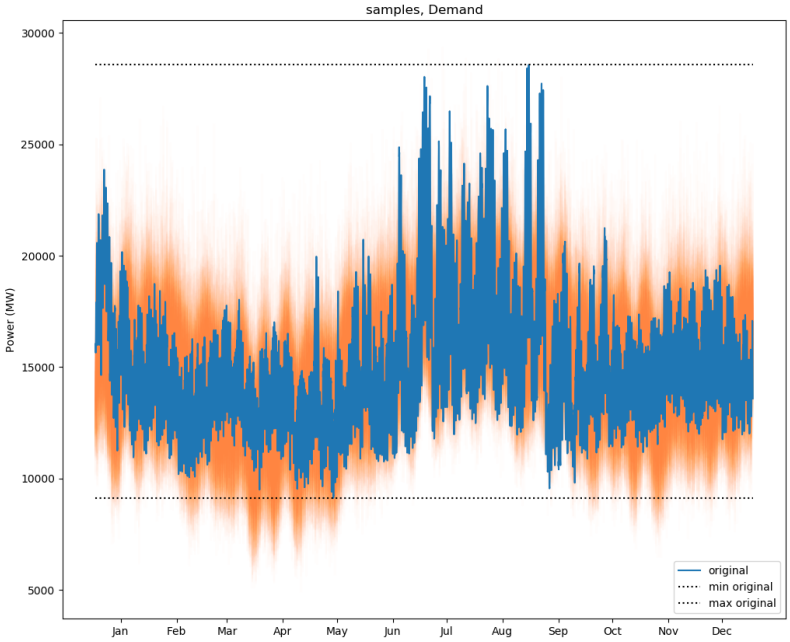

Figure 2-8. Training data with 100 VARMA synthetic samples.

This highlights the need to accurately capture the distribution of values from the training data reasonably, despite the time-dependent residual variance. Two variance handling methods for the VARMA were conceived and implemented as part of this work: segmentation and distribution preservation. These methods have both been added to the VARMA in RAVEN and are available as options to the user. Another requirement for the work presented here was a reduction in history length without reducing fidelity of the simulation. Due to the complexity of the economic market problem, some analyses cannot be performed on signals with high resolution because this easily results in hundreds of thousands of time steps to consider. On the other hand, smoothing out this resolution can destroy important physics involved in moment-to-moment operation of the grid components. As such, an ideal computational signal would comprise a small subset of a year-long history while both representing the full history as well as maintaining the fidelity of the signal. To deliver this, segment clustering was implemented in the VARMA algorithm in RAVEN.

The newly-developed algorithms of history segmentation, distribution preservation, and segment clustering are discussed in more detail in the following sections.

2.2.2.3.1 History Segmentation

Segmentation allows the training history to be divided into many discrete segments and have a separate VARMA model trained on each individually. In this manner, the seasonal time dependence of residual

variance can be localized and prevented from incorrectly informing other segments. The number of segments is left to the user.

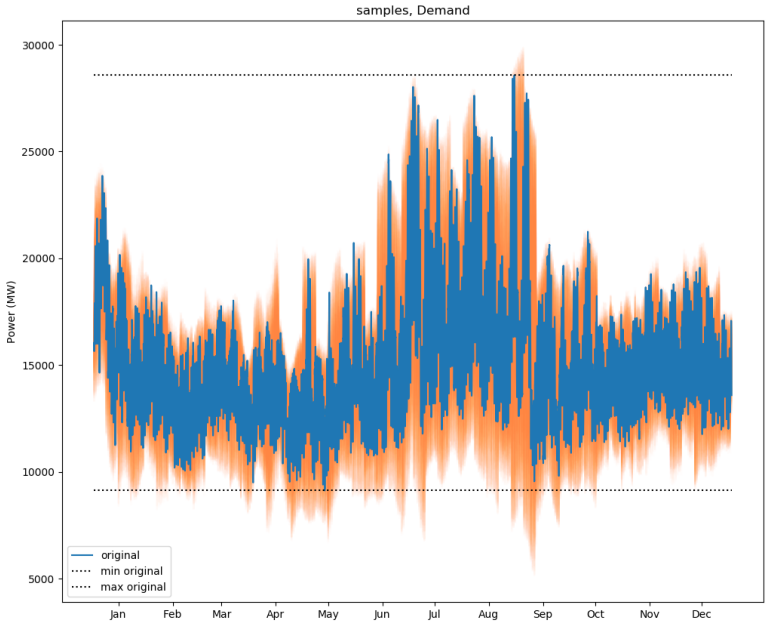

Figure 2-9. Training data with 100 VARMA synthetic samples, using weekly segmenting.

As shown in Figure 2-9, the abnormal lows are significantly reduced during the spring and fall, and the summer values are much more in line with the training data, although new lows are introduced around the seasonal transitions. This assures both peak and baseload power requirements will be more in line with the original training data. However, segmentation can have undesirable effects. Because each segment is represented exclusively by an independent model, abnormal events can only be replicated within that segment. For example, if an unusually high load is observed in the training data during February, and the training is segmented by months, the resulting variance introduced by the high load in February can only be seen in Februaries in the synthetic histories. For example, see the end of the summer peaks in Figure 2-9, where the high variance synthetic values are located immediately around the high variance in the original signal. Taken to an extreme, segmenting by days assures that a high load on a particular day in the training data yields a high load during the same day for all synthetic samples. This can lead to a synthetic sample that does not exhibit sufficient uniqueness or variation for the statistical analysis of a grid configuration. As a result, care should be taken to balance the benefits of segmentation in limiting time-dependent variance with the need for variation among samples.

2.2.2.3.2 Distribution Preservation

The second tool to address time-dependent signal variance is distribution preservation. This method seeks to nearly preserve the training data's statistical distribution in synthetic samples. By constructing an empirical distribution of both the training data and a given synthetic history, a transform can be applied to the synthetic history to restore the distribution of the training data in the synthetic sample. This approach preserves the capacity for particular measurements to be higher or lower in synthetic histories, but each

history will have the same general distribution as the training data; that is, the same total number of high and low values throughout the sampling space.

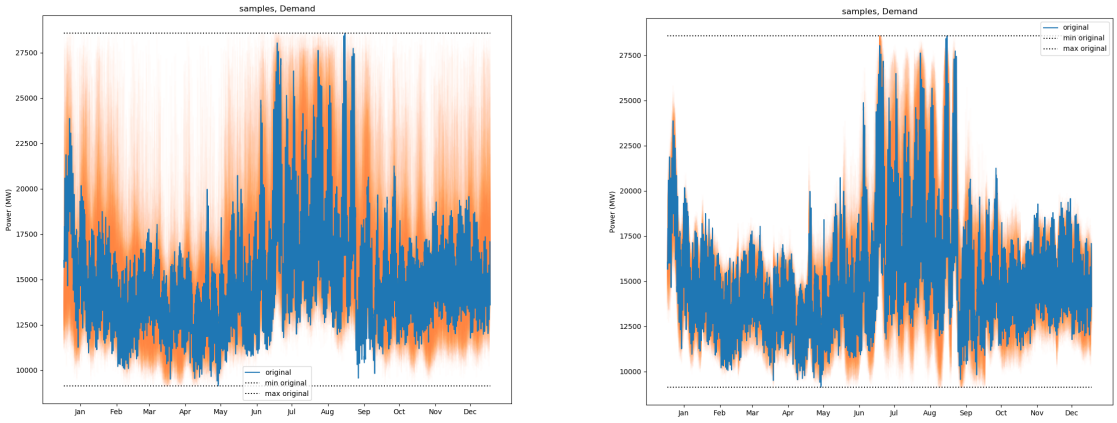

Figure 2-10. Training data with 100 VARMA synthetic samples (left) using distribution preservation, (right) using monthly segmentation and distribution preservation.

Figure 2-10 shows the use of distribution preservation both alone as well as along with segmentation. Note that without segmentation, while the distribution of the original training signal is preserved in the synthetic samples, occasionally very large loads are obtained in regions that are unrealistic. When both are used, there is still notable variability between synthetic histories without producing unrealistic signals.

The main observed drawback to distribution preservation is that the overall maximum and minimum values of the training data are enforced in the synthetic signal. This disallows unusual events to occur that would exceed the bounds of the original training signal. Future developments of this methodology may include relaxing the distribution matching, allowing some bounded variation while maintaining some of the original signal's distribution.

2.2.2.3.3 Segment Clustering

In order to preserve the fidelity of a high-resolution training set while also producing a history with a small number of points, a method of clustering history segments was devised. This method assumes that history segmentation is employed as discussed above, and as implemented is compatible with distribution preservation. The clustering concept is to observe identifying characteristics in the VARMA ROMs for each segment, and then group together VARMA ROMs whose representation is sufficiently similar. As a result, only one sample from each cluster is required to obtain a stochastic representation of the training data, possibly reducing the history length by a significant factor.

To perform segmented clustering training, first the Fourier detrending is divided into two groups: periods longer than the segment length, and periods within the segment length. Before segmenting occurs, the Fourier detrending is performed on all periods longer than the segment length and removed from the training signal. The partially-detrended signal is then segmented into user-defined lengths. For each segment, the remaining Fourier and VARMA training is applied to the associated section of the training data. The properties of each segment VARMA are then collected for segmenting. The clustering features for each VARMA include the shorter Fourier period coefficients λ_i and phase shifts η_i, the VARMA variance of the residual noise ϵ, and the mean of the long Fourier signal within the segment.

RAVEN provides access to a plethora of clustering algorithms, any of which can be applied to cluster the segment VARMA ROMs. Some clustering algorithms, such as *affinity propagation*, automatically determine the appropriate number of clusters to use based on the clustering features, while others such as

k-means take the number of clusters as input. The former is useful for determining the level of clustering reasonably available to the data, while the latter is useful for specifying the level of data reduction required.

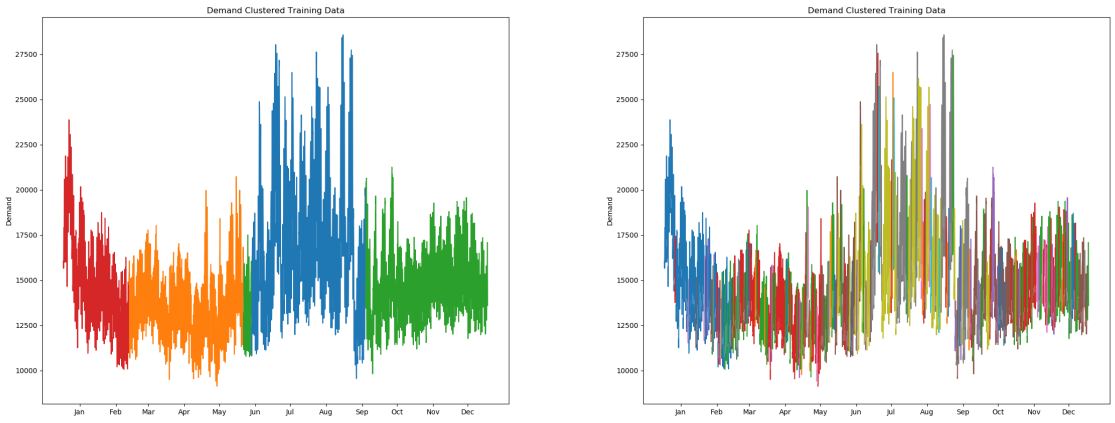

Figure 2-11. Clustering training data (left) weekly (right) daily.

After clustering the segments, the clustered VARMA is considered trained and prepared to sample. To sample the clustered VARMA, one random segment VARMA from each cluster is selected (including its long Fourier trend) and sampled. Finally, the resulting histories are placed together as a single history of disjointed segments. Note the cluster representation history (or *truncated history*) can be much smaller than the original history, and the segments are in no particular order. Rather, they represent the variety of prototypical signals that exist in the full synthetic histories, were they to be generated.

Note also when using the clustered VARMA that the truncated history does not represent the frequency of the clusters. That is, each cluster is represented exactly once, regardless of how many segments may belong the cluster. The multiplicity of each cluster is also made available so that analysis can be appropriately

extrapolated as though it were a full synthetic history. For example, see Figure 2-12. The representative samples are less than a tenth of the original signal, but demonstrate similar volatility, maxima, and minima.

The synthetic samples supplied to EDGAR in Section 3.2 for stochastic analysis of grid configuration employed all three of the improvements discussed, clustering on daily segments resulting in histories of 14 prototypical days.

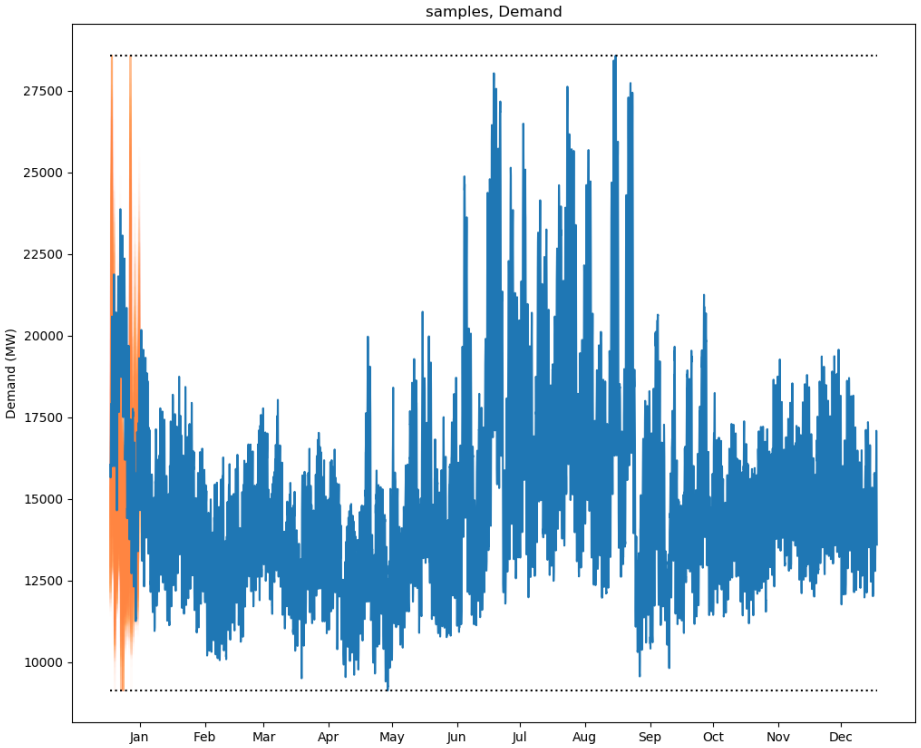

Figure 2-12. Truncated Synthetic Signals (orange) and original training data (blue) using 14 clusters of day-long segments arbitrarily located at the beginning of the year.

2.3 Development Pathway

The current daily market modeling capabilities developed and acquired by the SA&I Campaign can be used for performing a wide range of economic modeling analyses. For instance, one may look at modeling a specific U.S. grid market and to assess the impact of policies and technologies on the revenue of an individual unit. Alternatively, one may be interested at verifying the feasibility of a deployment scenario provided by GCAM or MARKAL, as discussed in Section 3 for demonstration purposes. Additional developments proposed in this section will enable these analyses and help building expertise on daily market economic modeling within the Campaign.

2.3.1 Improvements in EDGAR

Remaining effort in FY 2019 will focus on implementing storage unit modeling in EDGAR which requires some important modification of the economic dispatch logic. Currently, the ED optimization is performed for each hour, without possibility of anticipating on future hours. This implementation enables the smallest optimization problem and facilitates convergence to the best solution of the current ED problem since thermal units do not need to vary their power level in anticipation of a future higher or lower demand. However, storage units require to be able to anticipate future demand to decide if they need to store electricity or to wait until optimum time before releasing their stored energy. Consequently, the economic

dispatch optimization should be changed to enable optimization on multi-hour time-slices. This approach will also improve the current results obtained for nuclear load following, since those have physical constraints that would benefit from the capability to anticipate on future time-steps in order to decide if they should change their power output.

Additional effort is required to try to further reduce the computational burden needed to solve the UC problem by implementing simple machine learning strategies. The best UC solutions from the previous day can be provided on each new day as the first generation of UC solutions. This is not straightforward since the initial conditions for each unit will likely have changed and may require to re-sample some of them. Currently, the probability of undergoing different types of mutations is fixed, while it could be varied in the future from one generation to another in order to focus on the most efficient mutations.

Additional longer-term improvements are needed to extend the types of analyses done with EDGAR and should be the focus of future years. While EDGAR solves the day-ahead market problem, estimating the hour-ahead and real-time costs would enable to observe the short-term impacts of the uncertain prediction of renewable sources. The EDGAR code currently cannot estimate the unit revenue ensuing from the unit participation to ancillary service market. This problem is non-trivial as EDGAR uses a Monte Carlo approach, while the reserve income requires derivative calculations. Finally, continued improvement of the core structure, its computational performance, and the continued implementation of additional unit tests are required for quality assurance purposes.

2.3.2 Improvements in RAVEN/VARMA

2.3.2.1 Multiyear VARMA training

The current VARMA algorithm in RAVEN trains on a single time period and reproduces samples representative of that same time period. In the work described in Section 3.2, this means training on one year of data (2018) and producing synthetic histories that are representative of a single year. Often, more than one year of measurements are available for training the VARMA and could be used to improve the comprehension of local variance in the signal. This would allow several years of data to be provided as training materials.

In order to make use of multiple years in VARMA training, a few considerations need to be taken. For instance, consider grid load. When comparing historical data from 2008 and 2018 as training from the VARMA, for growing areas we expect the yearly mean load to increase. This increase is not generally cyclic, requiring special treatment to treat both years as indicative training data for general VARMA training. Identifying and removing growth trends could enable this use of multi-year data for VARMA training.

2.3.2.2 Physicality of VARMA clustering

The truncated histories produced as a result of VARMA clustering require detailed post-processing to determine the effectiveness of the mathematical clustering approach. The currently-implemented clustering technique is entirely mathematical and should be considered based on physical models to determine the validity of the clustering. For example, one might expect summer days to be clustered separately from winter days, and weekends and holidays separate from workdays. This validation effort would provide valuable insight into the clustering methodology and guide future developments in history reduction through clustering.

Similarly, outlier segments should be considered carefully when clustering, and further comparison to physical models is needed to determine how well these outliers are captured in the clustering algorithm.

For example, exploring how the clustering treats major holidays as exceptions to normal workday patterns should be considered to assure the appropriate variance of events is captured in the reduced histories.

2.3.3 Improve synergies between RAVEN and EDGAR

The current workflow to provide synthetic time series to EDGAR for grid configuration evaluation involves the following steps:

1. Obtain historical data for the desired time history trends (e.g. demand, prices, wind, solar, etc.).
2. Train VARMA models in RAVEN based on historical measurements.
3. Produce a series of synthetic signals using RAVEN as desired for use in EDGAR.
4. Deliver package of pre-sampled synthetic signals to EDGAR.
5. Perform EDGAR simulations on selected synthetic signals.

In other similar analyses [22, 24], pairing the grid dispatch optimization with the synthetic history production allowed an as-needed synthetic history production for the dispatch analysis, which reduces the manual workload and potential points of failure that comes from separating the two activities. In addition, the uncertainty quantification and optimization algorithms already existing in RAVEN become available for exploring various grid configurations in EDGAR via sampling a variety of VARMA time simulations. To restore this synergy, it is only necessary to create an interface between EDGAR and RAVEN. Because RAVEN is naturally designed to interface with other codes, the requirements to create an interface are well-established and understood and should require minimal effort. The resulting workflow would bring together the uncertainty quantification and synthetic sampling of RAVEN with the dispatch optimization of EDGAR, closing the gap that currently must be bridged manually.

In the approach proposed in Section 2.2.1.3.4, the profiles for the ancillary service requirements, i.e. regulation up and down, spinning reserve etc., were retrieved from load, wind power and solar power profiles by adopting suitable deterministic correlations. In common practice, the methods employed by the system operators to define operating reserve requirements are generally deterministic and depend only on the size of typical load variations [25]. They are insensitive to the level of renewable penetration in the system. Since deterministic approaches do not measure risk, in some circumstances complex high-risk situations are not sufficiently represented. Because of this, system operators are starting to transition from deterministic rules in favor of probabilistic methods to define their monthly reserve requirements [26]. From this standpoint, the RAVEN framework can be used to create a statistically-meaningful set of coherent scenarios using load and renewable patterns. By statistically approaching a particular energy system mix, reserve requirement profiles for EDGAR can be generated with RAVEN by means of a probabilistic approach. In this way, a trade-off between cost and risk can be quantified, instead of avoiding all risk at almost any cost.

3. Example of application: analysis of NY-ISO

This section illustrates the full method developed for grid analysis shown in Figure 1-1, where regional market analysis codes are used to drive simulations scenarios with daily market codes. It is applied for demonstration purposes and to build expertise on daily market analyses. The economic analysis of the New York region is performed to assess the impact of different fleet configurations on energy generation cost in 2050. The deployment scenario extracted from GCAM is used to provide long-term (2050) projection on the electrical grid components, those are being modeled using the EDGAR code to investiage the feasibility of this scenario based on hourly discretization. EDGAR relies on load and renewable data generated using VARMA. This exercise is especially useful to help the Campaign better understand what are the specificities of the different market modeling codes acquired and developed, what type of grid data is involved, what are the assumptions these codes rely on and their quantitative impact, and where are the remaining gaps in our tools those need to be addressed.

3.1 Deployment scenario description

Daily market codes use the installed capacity by technology provided by long-term market analysis models such as GCAM or MARKAL. The daily market analyses focus on a specific deregulated market. New York ISO (NY-ISO) has been selected for initial analysis, because it is a mostly deregulated market with significant fraction of nuclear and wind generations foreseen in 2050, and because the data available from NYISO can be readily used by RAVEN to generate representative synthetic samples as described in Section 3.2. Subsequent analysis will focus on Texas, and data from ERCOT will be used for training in that case. Either GCAM or MARKAL can provide the required information, with the following caveats:

- GCAM provides state-by-state electricity generation. The electricity generation can be converted into installed capacity by assuming an average capacity factor for each generation type (see Table 3-12).
- MARKAL projects the installed capacity for each of its 10 regions. The regions containing New York and Texas are respectively the Middle Atlantic (MDA) and West South Central (WSC) region. The MDA region, in addition to New York, comprises New Jersey and Pennsylvania. The WSC region includes Texas, Arkansas, Oklahoma and Louisiana. It is noted that Louisiana is a regulated state, but its overall contribution to the WSC region is sufficiently small that WSC might be treated as a deregulated region.

As noted in Section 2.1.1, the presented long-term market analysis is based on multiple scenarios that consider variations in demand, fuel costs, policy implementations, etc. Typically, a business-as-usual scenario is used as reference, where the current cost assumptions and policies are maintained for the duration of the simulation. In particular, reference scenarios generally do not model any explicit future climate policy interventions.

To demonstrate how the long-term market modeling tools interact with daily market analysis codes, an existing GCAM reference scenario was selected to provide boundary conditions (i.e. installed capacity and a consistent set of costs) to EDGAR. In this scenario, “business-as-usual” means that there is no explicit carbon tax in effect, but there are two relevant policies currently in place that may affect coal and other fossil fuels in the real world. The first is an emissions limit of 1,400 lbs of CO_2/MWh (which might be relaxed to 1,900 lbs CO_2/MWh in the near future). There is also a federal tax credit on CO_2 capture and storage, called Q45, that provides \$50/metric ton of CO_2 for sequestration. This is a credit and does not penalize fossil fuels. Other model assumptions include that current nuclear plants do not obtain a subsequent license renewal that would allow them to operate passed their current 60-year life, which results in the entire current fleet retiring by 2050. Retirements are complete for the nuclear plants in New York by 2040. New nuclear power plants are allowed after 2025 with a set of costs consistent with light water reactors and UO_2 fuel. The capital cost for new nuclear plants is 5,500 \$/KW, consistent with data published by the EIA

[27]. In this scenario, there is no technological learning accounted for, which should reduce costs and enhance nuclear penetration after the first few new builds.

The U.S. electricity generation grows by two thirds from the current 15 EJ to 25 EJ in 2050, but the electricity demand in New York is almost constant, because New York state is expected to be losing population in this time interval. The installed capacity shown in Figure 3-1 exhibits a slight increase because of the enhanced presence of variable renewables in 2050, with a low capacity factor as shown in Table 3-12. Coal and natural gas remain the main sources of electricity and comprise just over 50% of the total generation today, and just under 50% in 2050. Nuclear loses market share from around 18% in 2015 to 7% in 2050 because of plant retirements and limited builds. Variable renewables increase their penetration from around 5% in 2015 to 22% in 2050. Note that the scenario under consideration is policy neutral, and these behaviors are mostly driven by overnight capital costs.

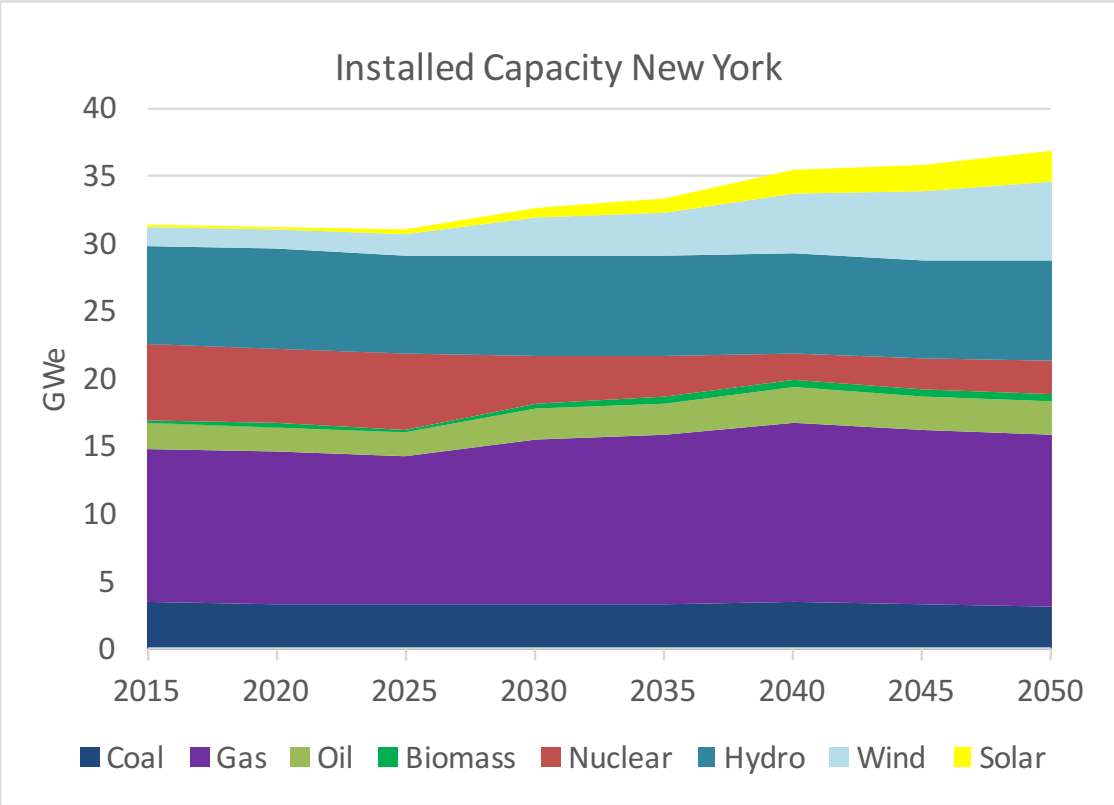

Figure 3-1. Installed capacity by generation type for New York (GCAM).

Table 3-1 shows the actual and GCAM electricity generation in 2015, the GCAM electricity generation in 2050, and the actual 2015 installed capacity. Electricity generation in 2015 looks consistent, with small differences between GCAM and actual data attributed to the fact that the model was calibrated to historical data until 2010 only, and not for 2015. The 2050 installed capacity used in the daily market analysis is scaled from the 2015 data according to the trends seen in the electricity generation. Work is underway to develop a capacity-dispatch version of GCAM that will eliminate this approximation, with variable capacity factors. Note that the conversion from generation to capacity is done in a post-processing step. The analysis performed with EDGAR in Section 3.3 can be used to confirm the feasibility of the modelled grid by showing that the demand and reserve requirements can always be met with installed capacity, and returning the effective capacity factors.

Finally, work is also underway to update the U.S. MARKAL multi-region model with a full set of consistent costs so that MARKAL results can in the future be used to inform daily market analysis.

Table 3-1. Actual and GCAM Electricity Generation and Installed Capacity.

	Generation (EJ/yr)			Installed Capacity (GWe)
	2015 (EIA)	**2015 (GCAM)**	**2050 (GCAM)**	**2015 (EIA)**
Coal	0.09	0.06	0.05	2.13
Gas	0.20	0.20	0.22	20.70
Oil	0.01	0.01	0.01	3.46
Biomass	0.01	0.00	0.01	0.57
Nuclear	0.16	0.16	0.07	5.40
Hydroelectric	0.09	0.09	0.09	4.7
Wind	0.01	0.02	0.07	1.75
Solar	0.00	0.00	0.01	0.08

3.2 Synthetic data generation

In order to consider the wide range of possible scenarios which a particular configuration of the NYISO grid is expected to cope with, synthetic samples are employed as described in this Section. Since EDGAR UC/ED simulations are computationally expensive, it was also critical to reduce the full year analysis to only a few representative days by using the clustering approach developed with VARMA. The generated samples are the result of training VARMA models within the RAVEN framework and sampling a number of times to produce the variable range of scenarios needed for EDGAR simulations.

To represent the New York region, historical data displayed in Figure 3-2 were obtained from the NYISO to provide total load (a.k.a. demand), wind electrical production, hydropower electrical production, and other renewable production (including biomass and solar) [28]. VARMA training is only performed on the total load demand and renewable generation. The other thermal generation will have their power output optimized with the EDGAR code. Data for the full year 2018 were retrieved from the NYISO website with

a five-minute resolution. These five-minute resolution data were averaged to produce hourly measurements of the four histories required by EDGAR for day-ahead UC/ED analyses.

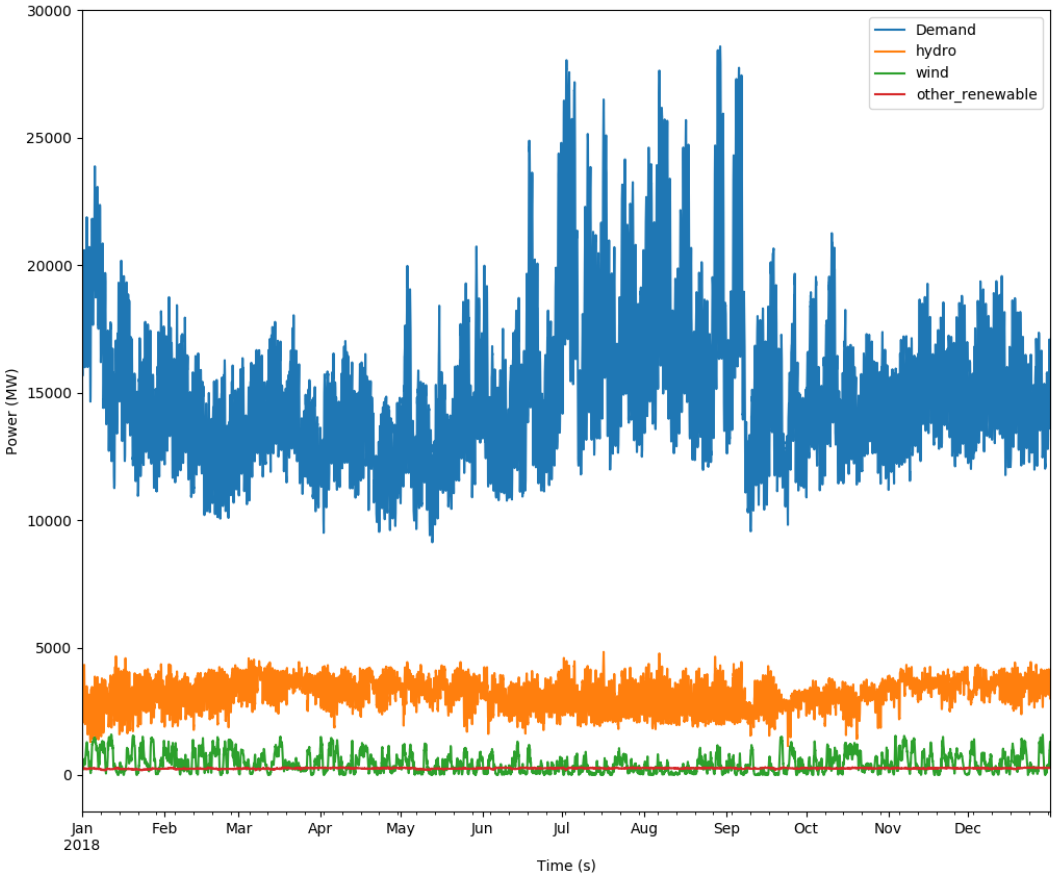

Figure 3-2. Hourly training data for 2018 NYISO.

The Fourier trends specific to each signal were determined using Fast Fourier Transform (FFT) analysis, and the desired cycles retained based on importance. The Fourier periods for each signal are displayed in Appendix A. Note that often the diurnal, yearly, and seasonal periods have significant impact on the individual signals.

After trying many potential combinations of clustering settings and segmentation lengths, clustering by days produced the reduced histories that most typified the original signal. The clustering strategy employed was *KMeans* with 14 clusters. The number of clusters was decided upon based on the recommendation of several other clustering strategies including *KMeansShift* and *AffinityPropogation*, and considering the limitations imposed by EDGAR's requested sample size. The resulting daily clusters appear in Figure 3-3. The different colors in Figure 3-3 indicate 14 clusters. Note that while Demand is shown here, the clustering algorithm is applied simultaneously to all four signals because they are assumed coupled and therefore non-separable.

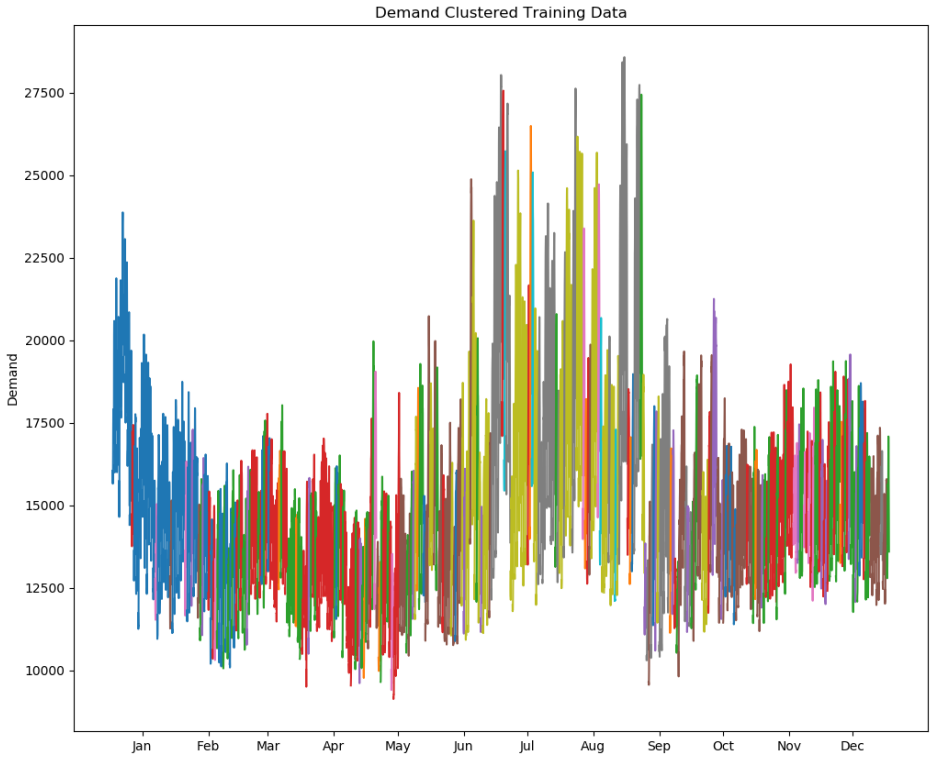

Figure 3-3. NYISO Daily Clustering, Demand.

To more clearly see where particular clusters appear, they can also be divided into individual plots as in Figure 3-4. Note that some clusters are seasonal (clusters 0, 2, 3, 7, and 8) while some are periodic. These clusters were obtained mathematically. Additional effort is needed to physically interpret what types of days are clustered (e.g. day of the week, season, working days, etc.). Again, only one history (demand) is shown by cluster here, but the clusters are identical among the metrics. The figure is intended to give a feel of cluster location, not the magnitude of each time history within each cluster.

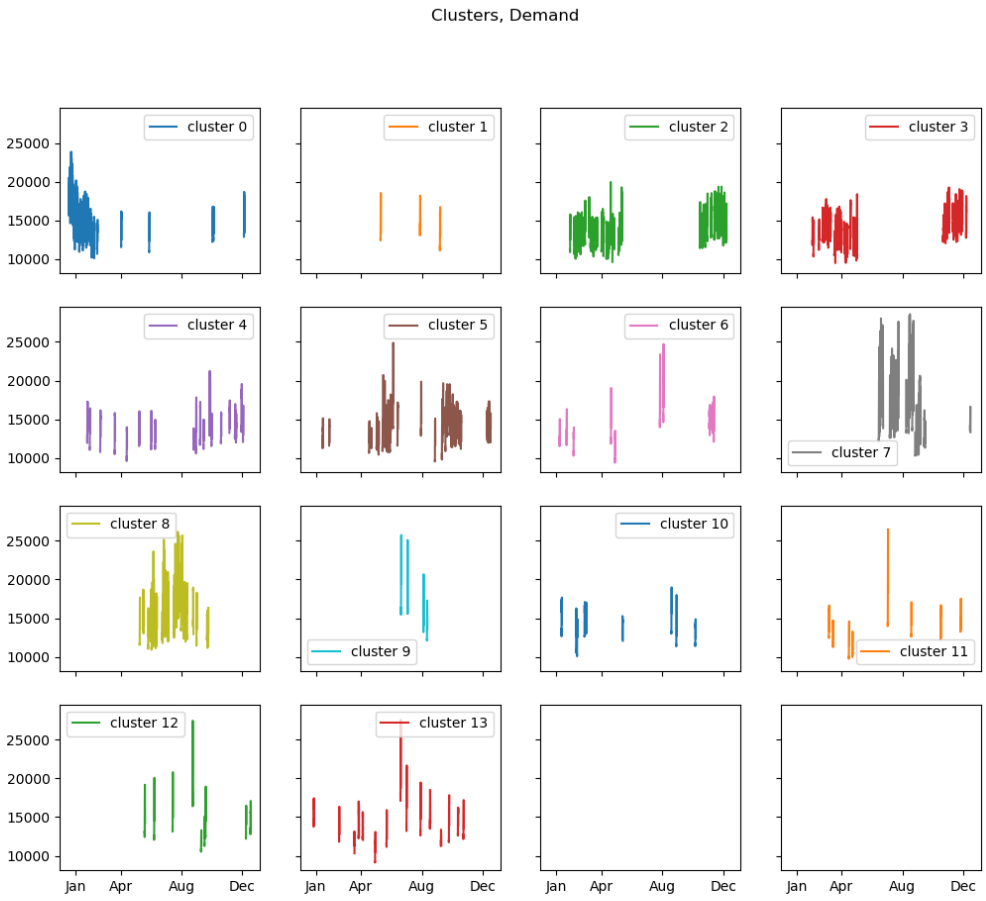

Figure 3-4. Daily Clustering by Cluster.

As in the process described above, 100 synthetic histories were produced by sampling one VARMA from each cluster and adding the associated Fourier trends. This yields a 14-day truncated history with each day comprising a single representative day from the year. Note also that the size of each cluster is different. Thus, the truncated history cannot be considered proportional to the year, but instead representative of the year. Weighting of the clusters is necessary to obtain a history proportional to the year using the weights shown in Table 3-2. The statistics of each cluster are available as a result of running the training workflow in RAVEN.

Table 3-2. Number of days represented by each day simulated.

Day	Number of days	Day	Number of days
1	46	8	37
2	3	9	45
3	51	10	4
4	54	11	9
5	21	12	8
6	51	13	9
7	12	14	15
		TOT	365

Finally, the resulting 100 synthetic histories are shown below for each of the four signals as a cloud of samples.

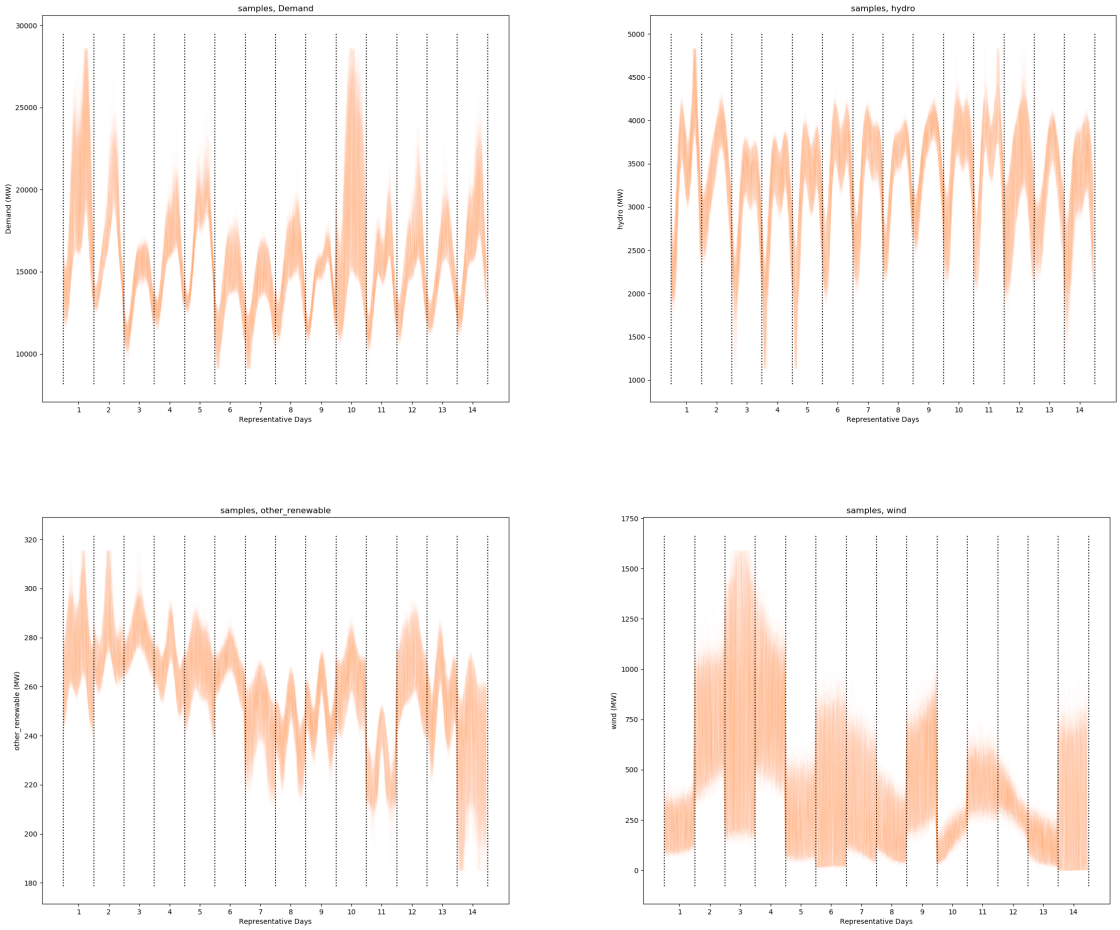

Figure 3-5. Synthetic Reduced Histories: Demand (top left), hydro (top right), other renewables (bottom left), wind (bottom right).

3.3 Daily market analysis

A set of UC/ED calculations was conducted with the scope of analyzing the NY-ISO market using the EDGAR code. The goal is to determine the optimal UC/ED schedule leading to the least cost of electricity, i.e. the different plant necessary to satisfy a given demand profile together with estimated reserve requirements. Simulations are first completed on the 2015 year using historic information from the grid, then on the 2050 scenario predicted from GCAM simulation.

3.3.1 Methods description and assumptions

- **Model for NY-ISO**

The first step is to build an EDGAR model of NY-ISO that contains detailed information on each operating power unit. This type of detailed information is not directly used by MARKAL or GCAM and needs to be retrieved from the ISO modeled. Special care should be taken to verify the consistency between the data retrieved, and the information provided by GCAM/MARKAL and VARMA. NY-ISO [29] contains data

regarding the generating facilities operating in the NY-ISO region in 2015. During the year, 697 units were operating in the region: 408 from renewable sources (biomass, hydroelectric, solar, wind) and 297 non-renewable sources (nuclear, coal, natural gas, oil, dual fuel) [29]. Table 3-3 shows the number of units and installed capacity per fuel type, while Table 3-4 shows the breakdown of natural gas units by technology type. The capacities of the different fuel generator types from [29] and shown in Table 3-3 present slight differences than those obtained through GCAM and shown in Table 3-1. The GCAM derived installed capacities were calculated by assuming the capacity factors shown in Table 3-12, while the one presented in [29] are the actual capacities in the region as presented by the system operator. The amount of electricity produced by the different fuel sources in 2015 in the NY-ISO region is reported in Table 3-5.

Table 3-3. Generating units in NY-ISO in 2015, extracted from NY-ISO [29].

Fuel type	**Number of units**	**Total installed Capacity (MW)**
Nuclear	6	5,440
Coal	8	1,469
Natural gas	54	4,086
Oil	84	3,091
Dual fuel (gas and oil)	145	19,283
Hydroelectric	347	4,267
Biomass, solar	46	507
Wind	21	1,461

Table 3-4. Natural gas units operating in NY-ISO in 2015, extracted from NY-ISO [29].

Fuel type	**Number of units**
Combined cycle	9
Combustion turbine	22
Jet engine	9
Steam turbine	6
Internal combustion	8

Table 3-5. Breakdown of NY-ISO electricity production in 2015.

	Energy (10^6 MWh)
Demand	134.58
Nuclear	43.06
Natural gas	26.40
Dual fuel	29.28
Other fossil	0.92
Hydroelectric	28.62
Wind	3.99
Other renewable	2.30

Each generating unit is characterized by a different capacity and operational parameters. The technologies can be categorized by their capability to follow rapid fluctuations of the load demands. Gonzalez-Salazar

et al [30] collected flexible operation features of different power generation technologies, which are reported in Table 3-6. Nuclear, lignite and coal power plants, certain steam turbines with oil/gas as boiler fuel are *inflexible* generation technologies. These power plants are designed for baseload operation, while start-up and ramping operations are rare and time-consuming. *Flexible* generation technologies comprise flexible gas turbine combined cycle, flexible coal, biomass, and biogas. These power stations are designed to adjust their generation level to cope with load variations and start at short notice. Reservoir hydro, combustion engines or aero-derivative gas turbines, a sub-set of simple cycle gas turbines are *highly flexible* technologies. The additional cost of operating these plants in a more flexible way can be very low.

Table 3-6. Flexibility characteristics of power generation technologies [30].

Technology	Minimum output (% full load)	Ramping rate (% full load/min)	Hot start-up time (h)
Hydro reservoir	5	15	0,1
Simple cycle gas turbine	15	20	0,16
Geothermal	15	5	1,5
Gas turbine (combined cycle)	20	8	2
Concentrated solar power	25	6	2,5
Steam plants (gas, oil)	30	7	3
Coal power	30	6	3
Bioenergy	50	8	3
Lignite	50	4	6
Nuclear	50	2	24

- **Description of the EDGAR simulation**

The goal of this analysis is to solve the UC/ED problem for one year, for the NY-ISO region. Simulating the whole year can be computationally overwhelming, and therefore a simplified approach was adopted. As discussed in Section 3.2, the UC/ED problem is solved on 14 nonconsecutive days selected by RAVEN/VARMA to be representative of the whole year. As a result, 14 UC/ED independent optimization problems are solved. However, in solving each day independently, the initial conditions that are chosen can affect the accuracy of the results.

In each of the 14 simulations, 2 consecutive days are simulated using the same demand and renewable production profiles. The first day is used to provide the initial conditions to the second day, which is optimized by minimizing the production cost. Once the results of each simulated day are obtained, the dispatch profiles and the generating costs are assumed to be the same for the number of days that the simulation represents and the power generated and the generating costs are multiplied by the respective number of representative days, as shown in Table 3-2. The resulting values representing the different sections of the year are then summed to obtain the generated electricity and generating costs for the whole year.

In this analysis, penalties associated with under-production and over-production of electricity are considered. A value of 3,500 $/MWh was chosen for both under-production and over-production scenarios. Penalties can be seen as a price to be paid to somebody else to absorb the electricity or to compensate for the missed demand. In reality, in an energy market there is no penalty associated with overproduction scenarios. However, in EDGAR the overproduction penalty serves the role to discard those non-optimal solutions and to encourage the system to meet the demand.

The electricity demand and the production profiles of wind, hydroelectric, and other renewables are treated stochastically. In each simulation, a demand profile is sampled through the VARMA algorithm along with the production profile of wind, hydro and other renewable sources. Then, the demand, “net” of renewable

production is calculated as the difference between the demand and the electricity produced through renewables. The net demand is dispatched through non-renewable components of the grid. Through this approach, there is no need to model the renewable units (biomass, hydroelectric, solar, wind), as their production profiles are sampled through the VARMA algorithm and imported as an input. Currently, the hydroelectric power is treated stochastically (the correlations between hydro-power generation and load demand are maintained with VARMA), while in a real case hydroelectric units vary their power output to respond to changes in demand or in renewable production. In addition, hydro-storage units are currently neglected in this analysis, while, in reality, these units play an important role in the grid. To this aim, hydro-storage is used to store energy when demand level is low, and to produce electricity during demand peaks. NY-ISO [28] shows a hydro storage installed capacity of 1.4 GW for 2015. Another important assumption to this model is that renewables do not contribute to meet reserve requirements.

- **Grid model simplification**

Through this approach, only the non-renewable units need to be modeled in EDGAR. However, modeling all 297 non-renewable units would require a high number of generations needed for convergence and be computationally expensive.

To this aim, gas, oil, and dual fuel units, which make up a total of 283 units, were assembled into 40 representative units. All units of the same fuel type were modeled with the same operational parameters while keeping the same total installed power. The number of units for each fuel type from Table 3-3, were reduced by a factor 283/40 and rounded to the nearest integers. For example, the installed capacity of natural gas units is 4,086 MW and it is produced by 54 plants, for an average of 76 MW/unit. In the simulation, the same installed capacity is obtained by 7 units of 583.7 MW. The representative units that were used in the simulation are shown in Table 3-7. Because the size of operating base-load units affects the reserve requirements, the number and capacity of nuclear and coal units were not modified.

Table 3-7. Representative units used in EDGAR simulations.

Fuel type	Number of units	Unit capacity (MW/unit)	Total Capacity (MW)
Nuclear	6	906.7	5,440
Coal	8	183.6	1,469
Natural gas	7	583.7	4,086
Oil	11	281.0	3,091
Dual fuel (gas and oil)	20	964.2	19,283

The normalized ramp rates, minimum and maximum output, and down times that were used for the different representative units are shown in Table 3-8. The values of the operational parameters used in the simulations, for the representative types of generators are shown in Table 3-9. The unit marginal costs were determined based on [2] and regional multipliers from NEMS. The no-load cost (NLC) represents the cost of maintaining the units on while producing the minimum output.

Table 3-8. Unit operational parameters per fuel type (normalized to the maximum output).

	Nuclear	Coal	Gas	Oil
Min output (MW)	100%	50%	50%	50%
Min down time (h)	8	8	3	3
Min Up time (h)	8	8	2	2
Ramp-down limit (% full load/h)	25%	25%	50%	50%
Ramp shutdown limit (% full load/h)	50%	50%	100%	50%
Ramp startup limit (% full load/h)	50%	50%	100%	100%
Ramp-up limit (% full load/h)	50%	22%	50%	100%
Max sustained ramp (% full load/min)	2.5%	2%	8%	8 %
Quick Start Capacity	0%	0%	70%	100%

Table 3-9. Unit operational parameters per fuel type.

	Abbreviation	Nuclear	Coal	Gas	Oil	Dual
Number of units		6	8	7	11	20
Marginal Cost ($/MWh)	MC	6.8	31.1	31.8	78.4	55.1
No Load Cost ($/h)	NLC	580	1,685	24,942	10,343	16,020
Startup cold cost ($)	SUCC	77,000	6,265	38,405	14,966	25,151
Startup hot cost ($)	SUHC	50,000	4,177	38,405	14,966	25,151
Shutdown cost ($)	SDC	0	0	0	0	0
Number of hours of a cold start, (h)	TCOLD	8	6	0	0	0
Max Output (MW)	PMAX	907	184	583.7	281.0	964.2
Min Output (MW)	PMIN	907	92	38	18	66
Min Down time (h)	MINDOWNTIME	8	8	3	3	3
Min Up time (h)	MINUPTIME	8	8	2	2	2
Ramp-down limit (MW/h)	RDL	227	45	292	141	482
Ramp shutdown limit (MW/h)	RSHUTDOWNLIM	227	92	584	141	964
Ramp startup limit (MW/h)	RSTARTUPLIM	227	92	584	281	964
Ramp-up limit (MW/h)	RUL	227	40	282	281	964
Max sustained ramp (MW/min)	MSR	23	4	47	22	77
Quick Start Capacity (MW)	QSC	0	0	378	412	964

3.3.2 Reference daily market analysis

3.3.2.1 2015

The electricity produced by each energy source in 2015 is shown in Figure 3-6 as reconstructed from the results of the 14 simulations. The reference data coming from NY-ISO [28] does not provide a breakdown between coal, dual-fuel, and gas plants, but an overall “fossil fuel” generation. The comparison between the electricity produced as evaluated by EDGAR and the reference data is shown in Figure 3-7. The amount of electricity produced by nuclear evaluated by EDGAR is 6.1% higher than the one observed in the region. The lower value in the “real” amount of electricity produced by is due to scheduled refueling and maintenance during 2015, which caused the shutdown of the units. At the moment, the maintenance

shutdown is not implemented in EDGAR. The amount of higher production by nuclear is counterbalanced by lower production of electricity from fossil fuels (-4.7%). The wind (+1%) and other renewables (+0%) production are consistent with the reference data, which confirms the synthetic data provided by VARMA are consistent with historic data.

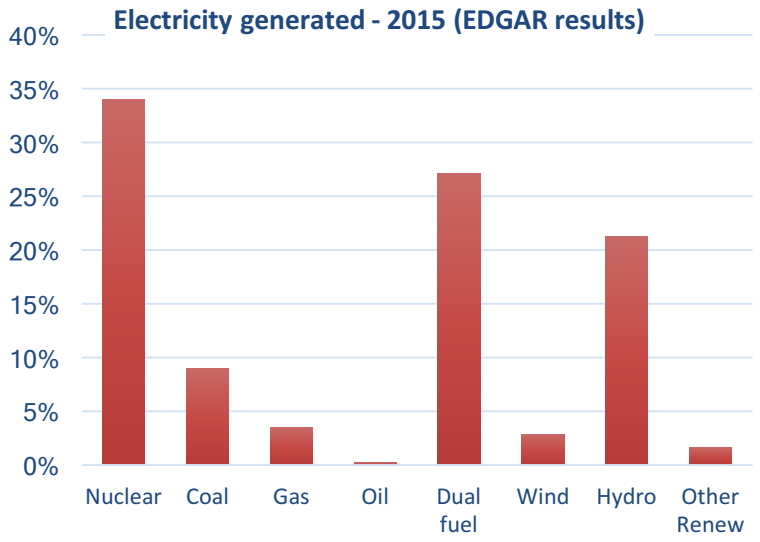

Figure 3-6. Breakdown of electricity generated by source calculated with EDGAR.

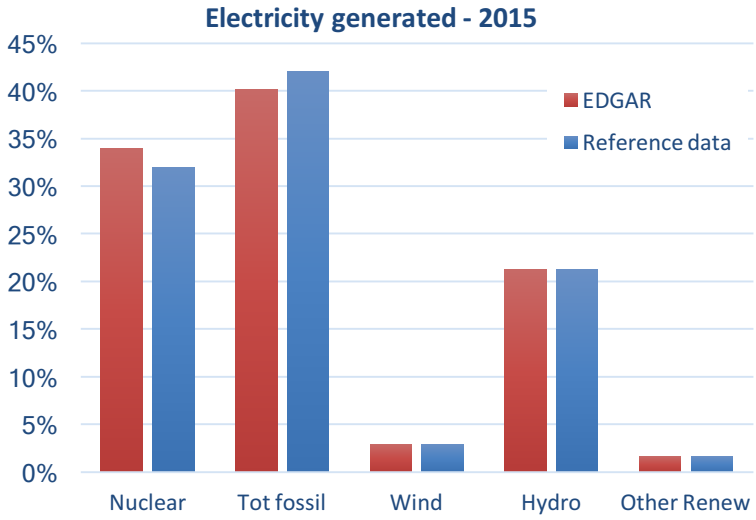

Figure 3-7. Comparison between EDGAR and reference electricity generated.

Using the first set of synthetic data, the electricity demand is matched at all times during the days, and the reconstructed total cost for the whole year is $4,180 M. The revenue per installed capacity from the nuclear reactors is 475.5 M$/GW installed, (57.1 $/MWh) while the revenue from the renewable sources is 2.0 B$ (57.4 $/MWh). The demand and production profiles as resulting from the first simulation of the first synthetic data are plotted in Figure 3-8, together with the generation cost. The clearing price for this

simulation is 78.36 $/MWh, which is equal to the marginal cost of the committed most expensive unit and is constant throughout this specific day.

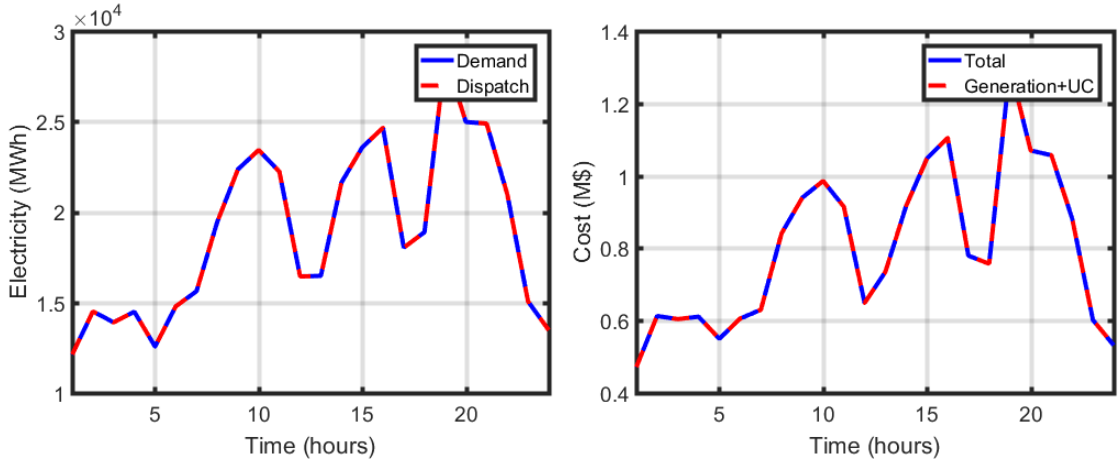

Figure 3-8. Demand and dispatch profile (a); total costs (b) (day 1 of 14).

The best cost as function of the generation number for the first simulated day (day 1 of 14) is shown in Figure 3-9. The plot shows that the calculation is converged after 26 generations.

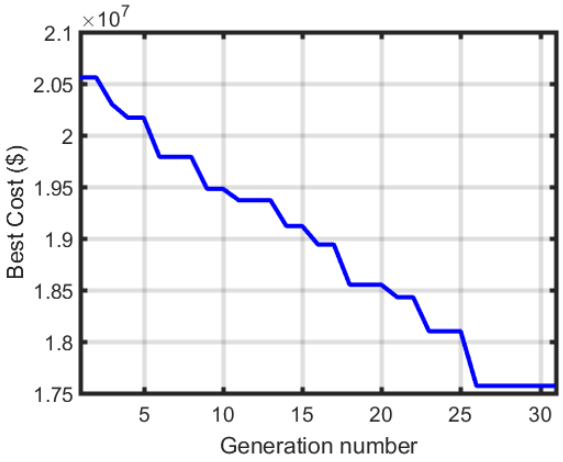

Figure 3-9. Best Cost as a function of the generation number (day 1 of 14).

- **Sensitivity on the number of representative units**

The first day of the first VARMA dataset (day 1 of 14) was also simulated with a different number of units to assess the impact of merging the 297 fossil-generator thermal units into 40. For the case presented in the previous section (with 40 units representing the fossil-generators), the total cost for the first day is $16.8

M. The UC cost is $0.3 M and the generation cost is $16.5 M. For this case, 26 generations are needed for convergence, and the total simulation time was 4 hours and 50 minutes on a 20-cores machine.

As the number of fossil units is reduced to 20, the total cost is $17.3 M. The UC cost is $0.2 M and the generation cost is $17.1 M. On the same 20-cores machine, the simulation time is reduced to 2 hours and 43 minutes, with 20 generations used for convergence.

The same day was run with 70 fossil units, increasing the number of UC scenarios from 100 to 300, and increasing the number of ED samples from 100 to 120. For this case, the total cost is $17.7 M, of which $17.4 M is generation cost and $0.3 M is UC cost. On an 80-cores machine, the simulation time was 9 hours and 32 minutes, which is equivalent to 38 hours and 8 minutes on a 20-cores machine. The best cost as a function of the generation number is shown in Figure 3-10.

This sensitivity analysis confirms the relatively low impact of the unit merging performed in this analysis (and further described in Appendix B) to reduce the computational burden.

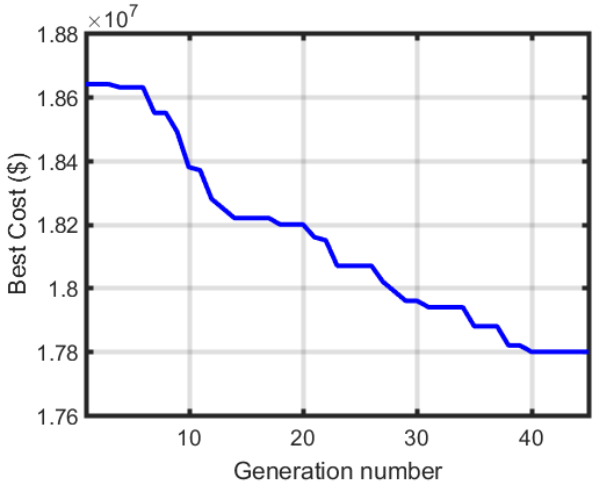

Figure 3-10. Best Cost as a function of the generation number (70 fossil units)

- **Sensitivity analysis on the synthetic data provided**

The EDGAR simulations were performed for 20 datasets of synthetic data for the 2015 scenario. The total cost resulting from the different VARMA datasets are shown in Figure 3-11. The averaged total cost is $4,228 M, with a standard deviation of $112.92 M, equal to 2.7% of the mean. The highest cost is obtained using the third dataset and is $ 4,293 M, 2.4% higher than the mean. These results show that the analysis for a single year can reasonably be performed using only the first set of data, which provides an average total cost. Additional sensitivity analysis on the set of synthetic data provided on the 2050 deployment scenario could not be performed for this report but would be required to assess the impact of the variations in synthetic data on the results obtained.

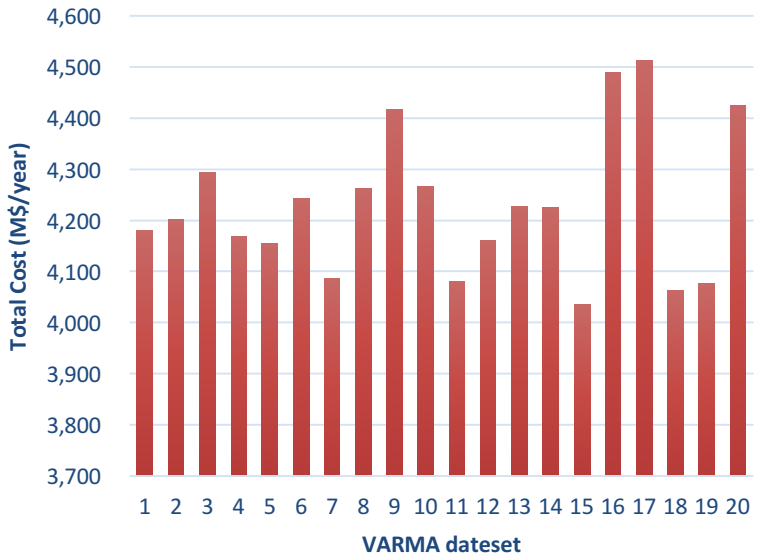

Figure 3-11. Total cost for different VARMA datasets.

3.3.2.2 2050

An UC/ED calculation was performed for year 2050, using inputs provided through the GCAM calculation presented in Figure 3-1. The relative percentage changes for the different energy sources are shown in Table 3-10. However, GCAM does not track plant capacity but only supplies energy generation. Therefore, the per unit change in electricity production were applied to the installed capacities in 2015 to estimate the installed capacities in 2050. Consequently, all unit types are assumed to have the same capacity factors in 2050 as in 2015. However, because of the increase of low capacity factor generators such as renewables (see Table 3-12), the overall installed capacity of the NYISO fleet increases by 21.7%, from 41.0 GW to 49.9 GW. Since GCAM does not provide the electricity generated by dual fuel plants, the increase in dual-fuel generating capacity was estimated as the average increase between the gas and oil capacity increases.

Table 3-10. GCAM electricity generation increase in the 2015-2050 timeframe.

Technology	2015-2050 increase	2050 Installed Capacity MW
Nuclear	-53.0%	2,559
Coal	-9.2%	1,334
Gas	+11.9%	4,561
Oil	+28.8%	3,980
Dual fuel[a]	+20.0%	23,176
Hydro	0.0%	4,267
Wind	+340.6%	6,436
Other Renewables	+330.5%	2,183
Total	+1.3%	49,905

[a] increase calculated as the average between the gas and oil increases

Table 3-11 shows the number of representative units used in the EDGAR simulations. The nuclear power was divided among two large plants of 907 MW (as the 2015 case) and three small plants of 248 MW. The total number of base-load units was limited to 40, as in the 2015 case, and the number was calculated according to the same methodology. For each of these sources, the total capacity was divided among the number of units.

Table 3-11. GCAM electricity generation increase in the 2015-2050 timeframe.

Technology	# units	Unit capacity (MW)
Nuclear (large)	2	907
Nuclear (small)	3	248
Coal	7	184
Gas	7	649
Oil	12	331
Dual fuel	21	1,102

According to the GCAM simulation, the electricity demand increases by 1.34% in the timeframe 2015-2050, and becomes 0.49 EJ (136,372 GWh). In year 2050, a 340.6% increase in wind generation and a 330.5% increase in other renewable generation (solar, biomass) are expected. The hydroelectric generation is constant over the period 2015-2050. To simulate the 2050 scenarios, the demand, wind, hydro, and other renewables profiles were increased using multiplicative factors representing the increases shown in Table 3-10. Out of consistency, the minimum power and ramp rates were re-calculated according to the same methodology used to simulate the year 2015.

Various scenarios were considered for the 2050 daily market analysis and are reported in Appendix C. Directly using the installed capacity extracted from GCAM, the 2050 results display significant over-capacity penalties, suggesting that the grid cannot always meet the demand at some hours due to large production from wind, hydro-power, and base-load nuclear. Renewable or nuclear curtailments prevent over-production penalties. Including hydro-storage or modeling hydro-generation and biomass as flexible generation units (instead of fixed generation) would provide different insight and should be considered in the future.

As a mean to avoid overproduction penalties, the 2050 case was simulated allowing renewable energy curtailment (RE-C). Curtailment is the reduction of output of a renewable resource below what it could have otherwise produced. The lower bounds of renewable production that were chosen are 80% for hydroelectric, 0 % for wind, and 0% for other renewables. In certain circumstances, curtailment allows a cost reduction, especially in those cases where certain units would be forced to turn off (with a subsequent shutdown cost) when renewable production is particularly high. In some cases, when the demand is low and the renewable production is high, system constraints might not allow certain units to shut down, with resulting over-production of electricity and subsequent penalties. In these cases, a reduction of the renewable power contribution might lower the penalty for over-production and the ensuing total cost.

The results show that over the whole year, wind production is curtailed by 14.3% ($0.57 \cdot 10^6$ MWh), other renewables are curtailed by 21.9% ($0.50 \cdot 10^6$ MWh), and hydroelectric is curtailed by 5.8% ($1.66 \cdot 10^6$ MWh). The load demand is met at all time and the total cost for the year is $ 4,469 M, 6.9% higher than the 2015 total cost. The revenue per installed capacity from nuclear sources is 490.9 M$/GW (56.1 $/MWh),

and the revenue from RE sources is 2.8 B$ (55.9 $/MWh), 3% and 40 % higher than those of 2015, respectively.

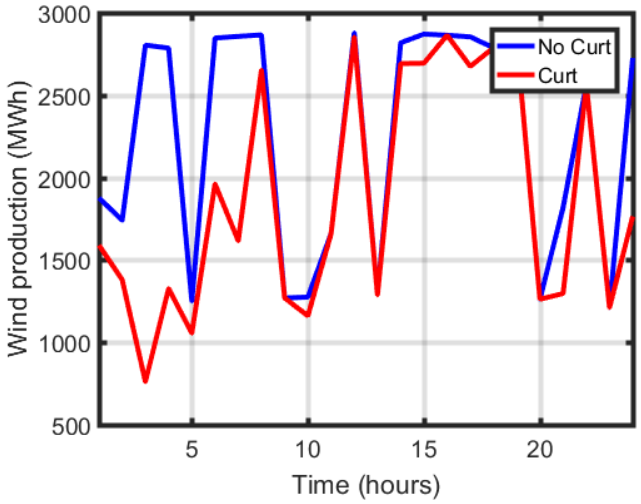

Figure 3-12. Electricity production by wind sources, without and with curtailment (day 11 of 14).

3.4 Summary of results

The analysis summarized in this section illustrates the full method developed for daily market grid analysis where a long-term scenario produced by GCAM for NY-ISO was used to drive simulations with the daily market analysis code EDGAR. The main outcome of this exercise is the expertise gained with completing this process through the following steps:

- Grid plant data need to be extracted from the grid modeled. Here we should in particular consider the number and capacity of each unit as the total installed capacity should be consistent between daily and regional market codes. Additional information needed by the daily market codes are the unit operational costs and capabilities. Simplification of the grid model (with reduced number of units) may be considered to speed-up the UC/ED computation without affecting the conclusion of the analysis.
- Cost data (Marginal Cost) and policy scenarios (e.g. carbon tax, …) should be consistent between daily and regional market codes for different capacity production technologies. This is fundamentally important since daily market codes prioritize production from lowest marginal cost technology, while regional market codes tend to prioritize installation of lowest capital cost technologies.
- Total load demand and renewable generation are extracted from the grid website and hourly data was generated using the VARMA module in RAVEN to condensate the full year statistical information within 14 representative days. Many sets of synthetic data were generated by VARMA and analyzed using EDGAR, showing relatively low variations in the total electricity generation costs (2.7% standard deviation and results spread over 5% of the mean value).

This approach enables to confirm the feasibility of the deployment scenario provided by capacity expansion code. It allows confirming the installed capacity is sufficient to meet the demand, that the VRE generation won't lead to exceed generation, that the VRE backup capacity requirements are sufficient, and that reserve requirements are met at every time of the year. The impact from the renewable generation curtailment

required on the reduction of the capacity factor (as shown in Table 3-12) is especially important and could be used back in the capacity expansion code.

This exercise highlights specific assumptions that are still used by different market analysis codes and helped quantify the impact of some of these. For instance, the unit merging approach developed provides a reasonable simplification to be used while additional effort focuses on reducing the computational burden from EDGAR simulations. However, the models for hydro power needs to be improved in EDGAR in order to address assumptions used to analyze a region like New York State. GCAM data is post-processed with constant capacity factors, but it could also be possible to use capacity factors from daily market simulations. The clustering approach employed with RAVEN's VARMA algorithm is still strongly user dependent and could be either data-informed or further automated in order to facilitate future analysis and ensure reproducibility.

Table 3-12. Generator capacity factors[a] by fuel type.

Technology	GCAM	EDGAR 2015	EDGAR 2050 RE-C
Nuclear	0.90	1.00	1.00
Coal	0.55	0.79	0.91
Gas	0.55	0.15	0.14
Oil	0.1	0.02	0.01
Dual fuel (gas-oil)	-	0.18	0.22
Pumped storage	-	0.00	0.00
Hydro	0.40	0.76 [b]	0.72 [c]
Wind	0.37	0.31 [b]	0.27 [c]
Other	0.55(biomass) 0.2(solar)	0.52 [b]	0.40 [c]

[a] The capacity factors are defined as the amount of electricity divided by the maximum amount that can be produced during the year. The maximum amount that can be produced during the year (in MWh) is the installed capacity (in MW) times the number of hours in a year (8,760).
[b] calculated in respect to the installed capacity shown in Table 3-3
[c] calculated in respect to the installed capacity shown in Table 3-10

4. SUMMARY

The System Analysis and Integration Campaign has been acquiring the capability and expertise to model energy market economics. The objective of this report is to describe the tools acquired for market analysis and to illustrate their capabilities and complementarities with an example of analysis.

Capacity expansion problems are solved with the GCAM and MARKAL models, simulating the energy markets in different U.S. and world regions over more than 100 years to provide a scenario-based long-term perspective. However, these simulations rely on simplified assumptions to account for the daily variations in electricity prices and unit generation, those need to be verified using smaller time-frames market analysis code systems. The EDGAR (Economic Dispatch Genetic AlgoRithms) code was developed under the SA&I to solve the combined Unit Commitment and Economic Dispatch problems to find the optimal schedule of a fleet of generating units to meet the forecasted grid demand over the next day in deregulated markets with an hourly time resolution. It can be used to demonstrate the feasibility of technology deployment scenarios proposed by capacity expansion codes. The capabilities of EDGAR were significantly expanded in FY 2018 & 2019, by improving its code structure and computation performance, adding physics modeling for xenon reactivity effect in nuclear reactors, optimizing the renewable curtailment, and enabling deterministic assessment of the reserve requirement. For these analyses, EDGAR relies on sets of load demand and renewable generation data with a one-hour time-step that are generated out of historic data using the VARMA (Vector Auto-Regressive Moving Average) model in RAVEN (Risk Analysis Virtual ENvironment), to condensate the full-year of data into a few representative days for EDGAR to simulate. To deliver this, segment clustering was implemented in the VARMA algorithm in FY 2019, together with two variance handling methods (segmentation and distribution preservation) conceived to better capture the distribution values from the training data.

This full suite of codes was used to model the New York ISO region in order to demonstrate the capabilities acquired and build expertise within the Campaign in daily market analysis. This exercise was especially useful to help better understand what are the specificities of the different market modeling codes acquired and developed, what are the assumptions these codes rely on, and where are the remaining gaps in our tools those need to be addressed. NY-ISO was selected because it is a mostly deregulated market with significant fraction of nuclear and wind generations foreseen in 2050. Long-term scenarios from capacity expansion codes can be used to drive simulations with daily market analysis codes following a few preliminary steps. First, consistent grid plant data and cost data must be gathered. Second, historic load demand and renewable generation data with fine time resolution should be obtained and processed through VARMA to condensate the full year statistical information into a few representative days. Daily market analyses are then performed with EDGAR to schedule a representative fleet of units (40 modeled for NY-ISO) on the reference time-point (2015) to demonstrate convergence of the results obtained, and their sensitivity to different sets of synthetic data generated with VARMA. Finally, similar analysis can be performed on the long-term (2050) scenario produced by GCAM to assess the feasibility of the deployment scenario provided. This procedure allows confirming the installed capacity is sufficient to meet the demand, which reserve requirements are met at every time of the year, and that the VRE generation won't lead to exceeding generation if their curtailment is allowed. The curtailment rate can then be provided back to the capacity expansion code in order to improve its model.

Consequently, the daily market analysis codes acquired by the SA&I Campaign enable analyses that are complementary to the global and regional energy market analyses. In particular, some of these codes are developed within the campaign as they provide unique capabilities for accurately modeling nuclear units and accounting for uncertainties in load and renewable generation data tools. Current applications discussed in this report are limited to testing and developing analysis experience while additional work is underway

to improve accuracy, performance, and to keep extending the type of analyses enabled with daily market modeling codes. Future efforts will also focus on applying this approach to additional U.S. regions.

REFERENCES

1. Nuclear Technology Research and Development. "Advanced Fuel Cycle Cost Basis – 2017 Edition," INL/EXT-17-43826

2. U.S. Energy Information Administration, "Annual Energy Outlook 2019"

3. S. Kim, A. Cuadra, B. Dixon, "Market Penetration of Nuclear Power under Various Technology and Climate Change Policy Scenarios", NRRD-FCO-2017-000056, 2017

4. A. Reisman, J. Lee, V. Bhatt, "A Market-based Methodology to Project Future Deployment of AFCI Technologies in the U.S. and Nuclear Energy in the World", Published as Appendix A, in Advanced Fuel Cycle Economic Tools, Algorithms, and Methodologies by D. Shropshire et al, INL/EXT-09-15483, 2009.

5. W. Banzhaf, P. Nordin, R.E. Keller, F.D. Francone, "Genetic Programming: An introduction", Morgan Kaufmann Publishers, Inc., San Francisco, California (1998).

6. N.P. Padhy, "Unit commitment – a bibliographical survey," IEEE Transactions on Power Systems, vol. 19, no. 2, pp. 1196-1205 (2004).

7. K.S. Swarup and S. Yamashiro, "Unit commitment solution methodology using genetic algorithm", IEEE Transactions on Power Systems, vol. 17, pp. 87-91 (2002).

8. N. Stauff, F. Ganda, E. Hoffman, T. Kim, S. Passerini, R. Ponciroli, T. Taiwo, B. Dixon, A. Epiney, C. Rabiti, P. Talbot, "Compatibility of Nuclear Technologies with Variable Grid Demand," NTRD-FCO-000263, September 1, (2017).

9. R. Ponciroli, Y. Wang, Z. Zhou, A. Botterud, J. Jenkins, R.B. Vilim, F. Ganda, "Profitability Evaluation of Load-Following Nuclear Units with Physics-Induced Operational Constraints", Nuclear Technology, 200(3), pp.189-207 (2017).

10. R. Fourer, D.M. Gay, B.W. Kernighan, "AMPL: A Modeling Language for Mathematical Programming", 2nd ed., Duxbury Press, Pacific Grove, California (2002).

11. B. T. Rearden, R. A. Lefebvre, A. B. Thompson, B. R. Langley, N.E. Stauff, "Introduction to the Nuclear Energy Advanced Modeling and Simulation Workbench," M&C 2017, Jeju Island, South Korea, April (2017).

12. Nicolas E. Stauff, Robert A. Lefebvre, Laura Swiler, Bradley T. Rearden, "Coupling of DAKOTA with the ARC suite of codes in the NEAMS Workbench for Uncertainty Quantification," ANS Summer meeting, Philadelphia, PA, USA, June 17-21, 2018.

13. Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.7 User's Manual.

14. Nicolas E. Stauff, R. Ponciroli, T. K. Kim, T. A. Taiwo, "Economic Impact of Flexible Nuclear Operation Estimated with EDGAR Optimization Code," proceedings of NURER2018, Jeju, Korea, 30 Sep. - 03 Oct. (2018).

15. "Nucléaire et Renouvelables, complémentaires pour réussir," special edition from the Revue Generale du Nucleaire (RGN), in French, Vol 1, Jan-Feb 2017.

16. L. Bird, J. Cochran, X. Wang, "Wind and Solar Energy Curtailment: Experience and Practices in the United States", NREL/TP-6A20-60983, 2014.

17. Z. Zhou, A. Botterud, J. Wang, R.J. Bessa, H. Keko, J. Sumaili, V. Miranda, "Application of probabilistic wind power forecasting in electricity markets", Wind Energy, 16, pp. 321–338, 2013.

18. M.A. Matos, R.J. Bessa, "Setting the Operating Reserve Using Probabilistic Wind Power Forecasts", IEEE Transactions on Power Systems, 26(2), pp. 594-603, 2011.

19. Z. Zhou, personal communication, Argonne National Laboratory, 2018.

20. G. Box, G. Jenkins, G. Reinsel, Time Series Analysis: Forecasting and Control. Prentice-Hall, (1994).

21. J. Chen, C. Rabiti, "Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems," Energy, 1-11(2016).

22. A. Epiney, C. Rabiti, P. Talbot, S. K. Jong, J. Richards, S. Bragg-Sitton, "Case Study: Nuclear-Renewable-Water Integration in Arizona," Idaho National Laboratory (2018).

23. C. Rabiti, "Status report on modeling and simulation capabilities for nuclear-renewable hybrid energy systems," Idaho Falls, ID: Idaho National Laboratory (2017).

24. C. Rabiti, A. Alfonsi, J. Cogliati, D. Mandelli, R. Kinoshita, "RAVEN, a New Software for Dynamic Risk Analysis," PSAM. Honolulu, Hawaii (2014).

25. M.A. Matos, R.J. Bessa, "Setting the Operating Reserve Using Probabilistic Wind Power Forecasts", IEEE Transactions on Power Systems, 26(2), pp. 594-603, 2011.

26. Z. Zhou, T. Levin, G. Conzelmann, "Survey of U.S. Ancillary Services Markets," Energy Systems Division, ANL/ESD-16/1, January 2016.

27. U.S. Energy Information Administration, Updated Capital Cost Estimates for Utility Scale Electricity Generating Plants, April 2013.

28. https://www.nyiso.com

29. NY-ISO, 2015. Gold Book – Load & Capacity Data, the New York Independent System Operator, Inc., 2015.

30. M. A. Gonzalez-Salazar, T. Kirsten, L. Prchlik, 2018. Review of the operational flexibility and emissions of gas- and coal-fired power plants in a future with growing renewables, Renewable and Sustainable Energy Reviews, Volume 82, Part 1, 2018, Pages 1497-1513

Appendix A: Fourier Frequencies from VARMA analysis

These tables are provided in the interest of repeatability as well as to provide guidance for future analyses.

Table A-1. Fourier frequencies, load.

Fourier Period (seconds)	Fourier Period (readable)	FFT Amplitude (W)
Infinite	**infinite**	1.35e8
86400	1 day	-8.26e6
1576800	6 months	5.26e6
31536000	1 year	-3.25e6
43200	12 hours	-2.66e6
1752000	20.27 days	1.87e6
955636	11.06 days	-1.81e6
3153600	36.5 days	1.72e6
1087448	12.6 days	-1.47e6
1314000	15.2 days	1.41e6

Table A-2. Fourier frequencies, hydro.

Fourier Period (seconds)	Fourier Period (readable)	FFT Amplitude (W)
Infinite	**infinite**	2.86e7
86400	1 day	-1.92e6
31536000	1 year	6.04e5
43200	12 hours	-5.93e5
7884000	3 months	-3.48e5
15768000	12 months	-2.37e5
10512000	4 months	-2.20e5
2866909	33 days	-2.19e5
769171	9 days	1.84e5
3504000	40 days	-1.84e5

Table A-3. Fourier frequencies, wind.

Fourier Period (seconds)	Fourier Period (readable)	FFT Amplitude (W)
infinite	**infinite**	3.99e6
31536000	1 year	7.53e5
1855059	21.5 days	-3.30e5
485169	5.6 days	2.76e5
342783	4 days	-2.74e5
618353	7.15 days	-2.20e5
685565	7.93 days	-2.20e5
2252571	26 days	2.08e5
1576800	18.25 days	1.98e5
700800	8.1 days	1.94e5

Table A-4. Fourier frequencies, other renewables.

Fourier Period (seconds)	Fourier Period (readable)	FFT Amplitude (W)
infinite	**infinite**	2.29e6
86400	1 day	-3.15e4
31536000	1 year	-2.59e4
7884000	3 months	1.94e4
43200	12 hours	1.87e4
2628000	1 month	1.80e4
3504000	40.5 days	1.56e4
630720	7.3 days	-1.40e4
1314000	15.2 days	1.34e4
1505143	52 days	1.16e4

Appendix B: Method Proposed for Merging Units in UC/ED Simulations

The sensitivity analysis performed in Section 3.3.2.1 revealed the importance of properly merging small units into larger one, in order to preserve the total generation cost together with the flexibility of the fleet.

In UC/ED simulations, the generation cost is calculated as:

$$CPG = NLC, \text{ if } P < P_{min} \iff P = 0 \quad (27)$$

$$CPG = NLC + \mathrm{MC} \cdot (P - P_{min}), \text{ if } P \geq P_{min} \quad (28)$$

where NLC is the no-load cost, MC is the marginal cost, P is the power level, and P_{min} is the minimum power of the unit.

When N units are assembled in a single unit, P_{min} is assumed to be the minimum power of the single unit to maintain the flexibility of the fleet. Consequently, the no-load cost of N assembled units (NLC_N) must be calculated as:

$$NLC_N = N \cdot NLC - MC \cdot (N - 1) \cdot P_{min} \quad (29)$$

Appendix C: Sensitivity daily market analyses on 2050 scenario

Various sensitivity analyses were performed with EDGAR on the 2050 scenario to find the most realistic grid operation that avoids penalties. The final scenario described in Section 3.3.2.2 considers curtailment of renewable generation but no nuclear load following.

- **2050 scenario without load following and renewable curtailment**

Year 2050 was simulated without allowing load following of nuclear plants and curtailment of renewable generation. The total cost for this scenario is $5,907 M, 41.2% higher than the 2015 total cost. The revenue per unit of capacity from the nuclear units is 538.9 M$/GW (61.5 $/MWh), while the revenue from renewables is 3.5 B$ (62.1 $/MWh). The high renewable penetration causes an overproduction above the energy demand at certain times of the day. Over the whole year, 0.3% of electricity is overproduced. However, because of the high unit-cost of missed demand, which is assumed here to be 3,500 $/MWh, the overproduction scenarios result in a yearly cost of $1,430 M, or 24.2% of the total cost. One of the main reasons for the high value of over-produced electricity lies on the assumptions made on the hydroelectric generators and storage. As hydroelectric is treated stochastically, there is no control on its production. In fact, hydroelectric is a highly flexible technology that allows to follow the demand and store energy when the demand exceeds the production, thus avoiding penalties.

- **Impact of nuclear load following**

The load following capability allows nuclear units to follow the demand and reduce the power output to find more optimal operation scenarios. The minimum power allowed for the nuclear units is 20% of the nominal power. For simplicity purposes, all nuclear units are supposed to be in their beginning of cycle and are not constrained by their xenon effect. This assumption will need to be revised in the future, but requires modification of the economic dispatch logic in order to allow it to anticipate on future time-steps before making a power change (as discussed in Section 2.3.1). Under these assumptions, the total yearly cost is $5,397 M, 8.5% lower than the nominal case. The overproduced electricity decreases from 0.3% of the nominal case to 0.1%, which translates to a penalty cost that is to 9.0% of the total cost. Over the whole year, nuclear production is reduced by 12% ($2.69 \cdot 10^6$ MWh) as compared to the case without load following. The revenue per GW installed from nuclear reactors is 442.2 M$/GW (60.7 $/MWh) and the

revenue from renewable sources is 3.5 B$ (62.9 $/MWh). With respect to the RE curtailment case, the revenue from RE sources is higher as more renewable energy is transmitted to the grid.

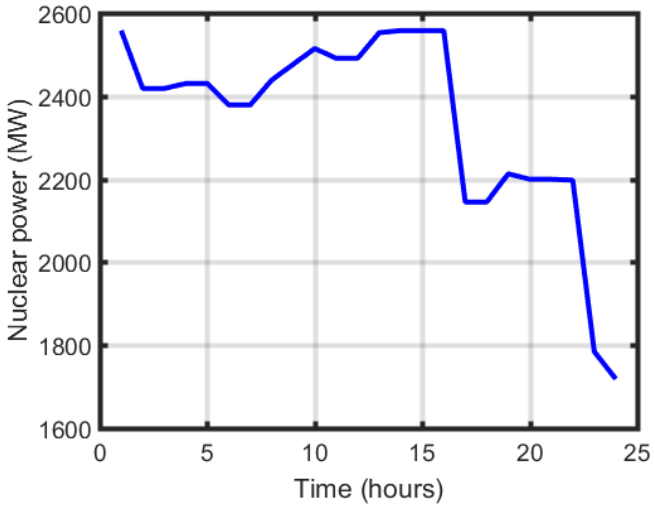

Figure C-1. Nuclear power production (total of all 6 nuclear units considered) for day 1 of 14, load following case.

- **Impact of curtailment and nuclear load following**

An additional case was run where the nuclear units were operated in flexible mode (load following) and the renewable power contribution could be curtailed. For this case, the total yearly cost is $4,648 M, 4% higher than the curtailment-only case (described in Section 3.3.2.2), and 13.9% lower than the load following-only case. The revenue from nuclear sources is 478.1 M$/GW (56.1 $/MWh), and the revenue from RE is 2.9 B$ (56.3 $/MWh). Over the whole year, nuclear is curtailed by 5.3% ($1.19 \cdot 10^6$ MWh), wind is curtailed by 13.0% ($0.52 \cdot 10^6$ MWh), other renewables are curtailed by 20.0% ($0.46 \cdot 10^6$ MWh), and hydroelectric is curtailed by 3.8% ($1.09 \cdot 10^6$ MWh). This case is equivalent to the curtailment-only case, with a constraint that is relieved, i.e. the power of the nuclear units is not fixed at maximum power at all times. Therefore, it might sound surprising that the same optimization problem, with some constraints relieved, gives a less optimal solution. However, when load following is modeled, an additional constraint is added to the problem. When a nuclear unit reduced its power output, nominal operating conditions at rate power cannot be restored right after. The unit is constrained to be operated at reduced power output for two consecutive hours. The energy dispatch algorithm optimizes the dispatch in an hour-by-hour case, finding the configuration for the following hour that minimizes the total cost. Therefore, it may happen that the optimal solution for a given hour is driven by having the nuclear units at a low power level. However, in this case, the algorithm does not consider that for the following hour, more expensive units (e.g. gas, oil) need to be activated if the net demand increases. This case can be seen during day 14 at hour 20, when the second

nuclear unit is decreasing in power (Figure C-2). At the following hour (hour 21), the nuclear unit is forced to "stay down", and a gas unit (unit #3) increases power in to satisfy the demand.

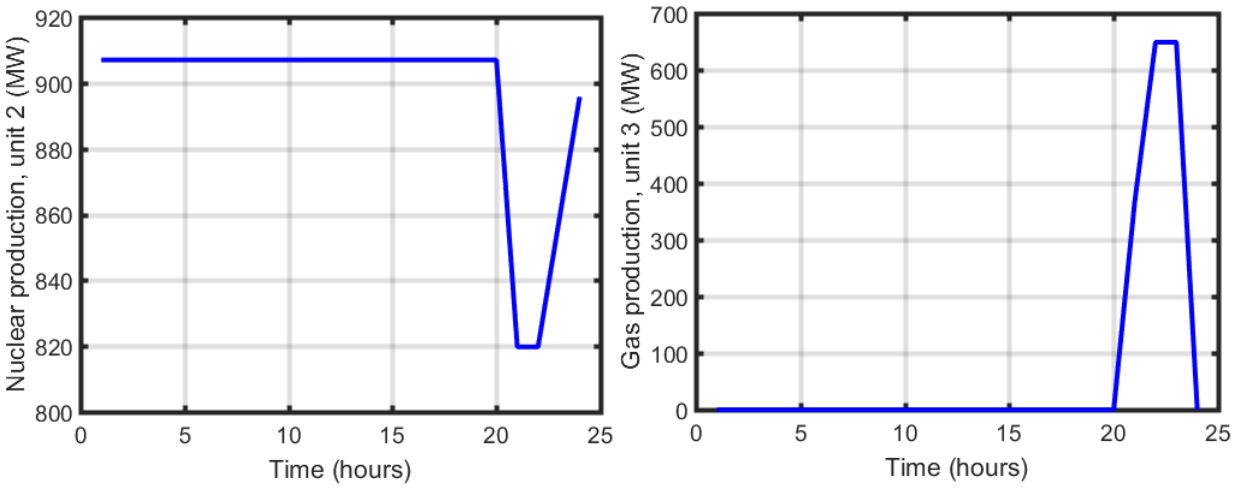

Figure C-2. Comparison between nuclear (left) and gas (right) production, with load following and curtailment.

This non-optimal scenario can be discarded if the energy dispatch is optimized on a time scale of several hours and not on an hour-by-hour base. This type of constraint typical to nuclear and to storage units illustrates the need for the economic dispatch to be able to anticipate on future requirements, as already discussed in Section 2.3.1.

- **Sensitivity analysis on the nuclear capacity**

A sensitivity analysis was conducted on the installed capacity of nuclear, considering load following and curtailment. With respect to the case described in the previous section, a large nuclear unit is added to the fleet. The total nuclear capacity is 3,465 MW, i.e., 35.5% higher than the reference case (2,558 MW installed). This scenario leads to a total cost of $4,264 M, 8.3% lower than the case without the additional nuclear unit. Over the whole year, nuclear is curtailed by 5.0%, ($1.12 \cdot 10^6$ MWh), wind is curtailed by 16.9% ($0.67 \cdot 10^6$ MWh), other renewables are curtailed by 23.4% ($0.54 \cdot 10^6$ MWh), and hydroelectric is curtailed by 5.3% ($1.52 \cdot 10^6$ MWh). As nuclear capacity is added, the total cost decreases, but more renewable production is curtailed in order to avoid over-production scenarios. The revenue from nuclear reactors is 453.4 M$/GW (55.2 $/MWh) and the revenue from renewable sources is 2.8 B$ (55.2 $/MWh). Compared to the previous case, the revenue from nuclear is 7.6% lower.

- **Summary of 2050 results**

The results of the sensitivity analyses are summarized in Table C-1. The highest cost ($5,907 M) is for the scenario without both renewable curtailment and load following, as it is characterized by a consistent overproduction penalty. As the cost of overproduced electricity does not have a realistic meaning the cost of this scenario is not realistic.

As RE-C is introduced, the overproduction penalty decreases to 0$ and the total cost decreases to $4,469 M. The total cost for the load following case ($ 5,397 M) is 20.7% higher than that of RE-C. The higher cost is due to the cost of electricity overproduced when renewables production is high.

As both RE-C and LF are used, the cost is slightly higher than the case with RE-C only ($4,648 M compared to $4,469 M). As previously explained, the reason for this lies in the nuclear reactor constraint to not increase power for 2 hours after a power decrease.

As additional nuclear capacity is installed, the cost decreases from $ 4,648 to $4,264 M, considering both RE-C and LF.

The revenues from nuclear and RE do not change drastically across the scenarios. The higher revenues for the case without RE-C and LF are mainly due to the higher amounts of electricity produced and do not account for the energy overproduced.

Table C-1. Summary of sensitivity results.

Scenario	Total cost (M$)	Nuclear revenue (M$/GW)	RE revenue (B$)
RE curtailment (RE-C)	4,469	490.9	2.8
Nuclear load following (LF), no RE-C	5,397	442.2	3.5
No RE-C, no LF	5,907	538.9	3.5
RE-C and LF	4,648	478.1	2.9
Additional nuclear, RE-C and LF	4,264	453.4	2.8