



# A Hybrid Energy System Workflow for Energy Portfolio Optimization

July 2021

*Changing the World's Energy Future*

Jia Zhou, Paul W Talbot, Cristian Rabiti, Hany Abdel-Khalik



*INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC*

#### **DISCLAIMER**

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

# **A Hybrid Energy System Workflow for Energy Portfolio Optimization**

**Jia Zhou, Paul W Talbot, Cristian Rabiti, Hany Abdel-Khalik**

**July 2021**



**Idaho National Laboratory  
Idaho Falls, Idaho 83415**

**<http://www.inl.gov>**

**Prepared for the  
U.S. Department of Energy  
Under DOE Idaho Operations Office  
Contract DE-AC07-05ID14517**

## Article

# A Hybrid Energy System Workflow for Energy Portfolio Optimization

Jia Zhou <sup>1,\*</sup> , Hany Abdel-Khalik <sup>1</sup>, Paul Talbot <sup>2</sup>  and Cristian Rabiti <sup>2</sup><sup>1</sup> School of Nuclear Engineering, Purdue University, West Lafayette, IN 47906, USA; abdelkhalik@purdue.edu<sup>2</sup> Idaho National Laboratory, Idaho Falls, ID 83415, USA; paul.talbot@inl.gov (P.T.); cristian.rabiti@inl.gov (C.R.)

\* Correspondence: zhou646@purdue.edu; Tel.: +1-765-775-3075

**Abstract:** This manuscript develops a workflow, driven by data analytics algorithms, to support the optimization of the economic performance of an Integrated Energy System. The goal is to determine the optimum mix of capacities from a set of different energy producers (e.g., nuclear, gas, wind and solar). A stochastic-based optimizer is employed, based on Gaussian Process Modeling, which requires numerous samples for its training. Each sample represents a time series describing the demand, load, or other operational and economic profiles for various types of energy producers. These samples are synthetically generated using a reduced order modeling algorithm that reads a limited set of historical data, such as demand and load data from past years. Numerous data analysis methods are employed to construct the reduced order models, including, for example, the Auto Regressive Moving Average, Fourier series decomposition, and the peak detection algorithm. All these algorithms are designed to detrend the data and extract features that can be employed to generate synthetic time histories that preserve the statistical properties of the original limited historical data. The optimization cost function is based on an economic model that assesses the effective cost of energy based on two figures of merit: the specific cash flow stream for each energy producer and the total Net Present Value. An initial guess for the optimal capacities is obtained using the screening curve method. The results of the Gaussian Process model-based optimization are assessed using an exhaustive Monte Carlo search, with the results indicating reasonable optimization results. The workflow has been implemented inside the Idaho National Laboratory's Risk Analysis and Virtual Environment (RAVEN) framework. The main contribution of this study addresses several challenges in the current optimization methods of the energy portfolios in IES: First, the feasibility of generating the synthetic time series of the periodic peak data; Second, the computational burden of the conventional stochastic optimization of the energy portfolio, associated with the need for repeated executions of system models; Third, the inadequacies of previous studies in terms of the comparisons of the impact of the economic parameters. The proposed workflow can provide a scientifically defensible strategy to support decision-making in the electricity market and to help energy distributors develop a better understanding of the performance of integrated energy systems.



**Citation:** Zhou, J.; Abdel-Khalik, H.; Talbot, P.; Rabiti, C. A Hybrid Energy System Workflow for Energy Portfolio Optimization. *Energies* **2021**, *14*, 4392. <https://doi.org/10.3390/en14154392>

Academic Editor: Valerio Lo Brano

Received: 6 April 2021

Accepted: 9 July 2021

Published: 21 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Keywords:** economic optimization; Gaussian process; synthetic time history generation; integrated energy systems; HERON; RAVEN



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

To optimize energy and utilization configurations, the U.S. Department of Energy (DOE) Office of Nuclear Energy (NE) established the Integrated Energy Systems (IES) program. This program's aim is to adopt innovative solutions to system integration and process design [1] by increasing the use of resources, energy efficiency, and system reliability. The IES model also takes into account all available energy sources to optimize its benefits while minimizing its less desirable qualities [2]. In 2016, natural gas replaced coal as the

most commonly used fuel in the United States to generate electricity, and it is projected to remain the leading source of electricity [3].

On the other hand, traditional baseload production has experienced a dramatic downturn in the energy market as Variable Renewable Energy (VRE) sources are benefiting from their low marginal cost. According to [4], in 2017, U.S. wind capacity increased by more than 8.3%, while solar capacity increased by 26% compared to 2016, accounting for more than 54% of newly installed renewable electricity capacity in 2017. This growing penetration of renewable energies will have unexpected impacts on the economic feasibility of traditional baseload technologies in the U.S. and around the world.

The first impact is on the price of electricity. When renewable energy production is higher than demand, the price could be negative [5]. This creates another impact on the traditional baseload energy producers—currently, they must either limit their production or waste power. To maintain economic competitiveness in this changing market, many U.S. nuclear power plants are now beginning to assess the technical and economic feasibility of redirecting surplus energy to other services [6]. Several studies show that supporting current nuclear plants is cost effective in minimizing CO<sub>2</sub> pollution; however, nuclear zero-emission combustion is not respected by deregulated markets. It is also reported that the energy market has decreased nuclear plant income from energy purchases while also raising operational and maintenance costs [7]. Another study also highlights that two-thirds of the U.S. nuclear capability is unprofitable and one-fifth of nuclear power plants are likely to retire early, with inexpensive gas as a key catalyst for nuclear productivity failure. The premature closure of some nuclear power plants is motivated by these economic difficulties, driven by both the growing renewable and low-cost natural gas in the U.S. [8]. Moreover, the increasing penetration of renewable energies increases net load volatility, where the net load is the difference between the total electric demand and the renewable portion. That volatility must be balanced by other sources. Traditional energy producers have started to meet the evolving grid conditions by varying their production. Growing fluctuations in net load have been shown to require generator flexibility on all time scales and various spatial scales [9]. On the other hand, energy storage and the fuel cell industry [10] are critical to reducing the size and operating cost of hybrid energy systems [11].

A new type of energy system is needed to minimize the overall system cost and maximize the usage of different resources to increase the system's reliability. It has become increasingly complex to select an appropriate energy portfolio for a particular market. It is not a straightforward choice because utilizing one energy option may affect the selection of other energy options. Decision-makers need methods and tools for evaluating whether an energy portfolio will lead to reliable service at reasonable rates and follow the CO<sub>2</sub> emission regulations. Besides, there are new problems that arise in determining the value of different components in an energy portfolio. For instance, the cost of building and operating VRE sources is relatively low. However, the availability of VRE is highly time-dependent, and the available hour-to-hour or day-to-day VRE quantity is rather difficult to predict, given its sensitivity to climate conditions. The unreliable VRE supply will decrease the reliability of the electricity grid and end up adding the cost of the installation of flexible energy producers. This implies the need to obtain a holistic understanding of the value of various components in an energy portfolio. Thus, an optimization framework for the mixed-energy production portfolio is required.

Some commercial software can be used to optimize a mixed-energy production portfolio. However, the large volume of data input and long computing times add complexity to end-users. To combat this challenge, the Idaho National Laboratory's (INL) RAVEN framework has been employed to develop and implement a theoretical basis for the IES techno-economic analysis [12]. This analysis evaluates the technological and economic efficiencies of a process, product, or service. It typically incorporates process modeling, engineering design and economic assessment [13]. An economic analysis of the viability of nuclear power plant retrofitting has been conducted in [14]. The application of RAVEN in a case study that maps physical performance onto economic performance is reported in [15]. The financial performance of the RAVEN framework is further evaluated

by Reference [16]; the results show that economic benefits can be achieved by adding a suitable industrial process to the base-load generator. These past works, however, have relied on restrictive workflows, requiring complex operations that are not easily accessible to end-users, limiting their use to the advanced RAVEN users. To overcome these challenges, the HERON (Holistic Energy Resource Optimization Network) plugin was recently developed for energy dispatch optimization for energy system analysts [17]. HERON automates the construction of RAVEN-input templates suitable for solving the energy dispatch optimization problem. In tandem with HERON, another plugin TEAL (Tool for Economic AnaLysis) is used to perform the economic analysis.

This work develops a new optimization workflow integrated under HERON to optimize the size of installed capacities for different energy-producing units, including both renewable and conventional baseload energy producers. The goal of the optimization is to minimize the overall cost of energy production required to meet the energy demand, taking into account the seasonal demand variations and associated uncertainties, the construction and operational costs of the various energy units, as well as techno-economic factors such as discount rate, depreciation rate, inflation and taxes, and so forth. The next section discusses the various components of the new workflow.

## 2. Calculation Flow

The overall calculation flow of the optimization process is shown in Figure 1. It may be divided into three steps:

- The first step, depicted in a flowchart in Figure 2, reads the available historical data and builds the energy demand and generation models; details are given in Section 3.
- The second step, representing the bulk of the work automated by HERON, generates synthetic time histories using reduced order modeling (ROM), to be evaluated by HERON and TEAL; details are given in Section 4.
- The last step employs a Gaussian-Process-based model to optimize the overall cost of energy production; details are given in Section 5.

Synthetic time history generation is employed to expand the samples in our system with several sub-steps, including segmentation and clustering, Fourier detrending, Autoregressive Moving Average (ARMA) and peak detection. The synthetic time histories trained from this step are shown in a blue dashed-line box in Figure 1.

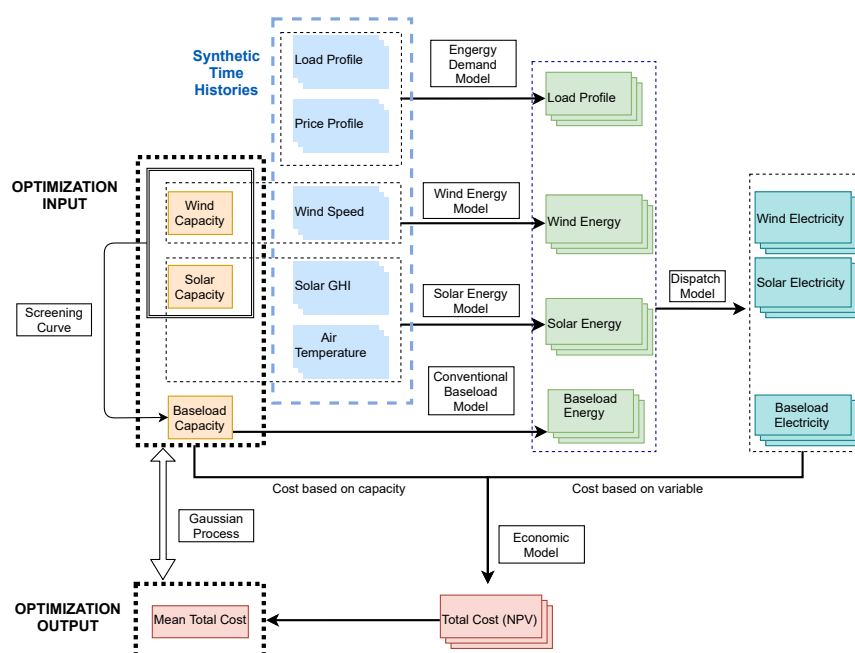
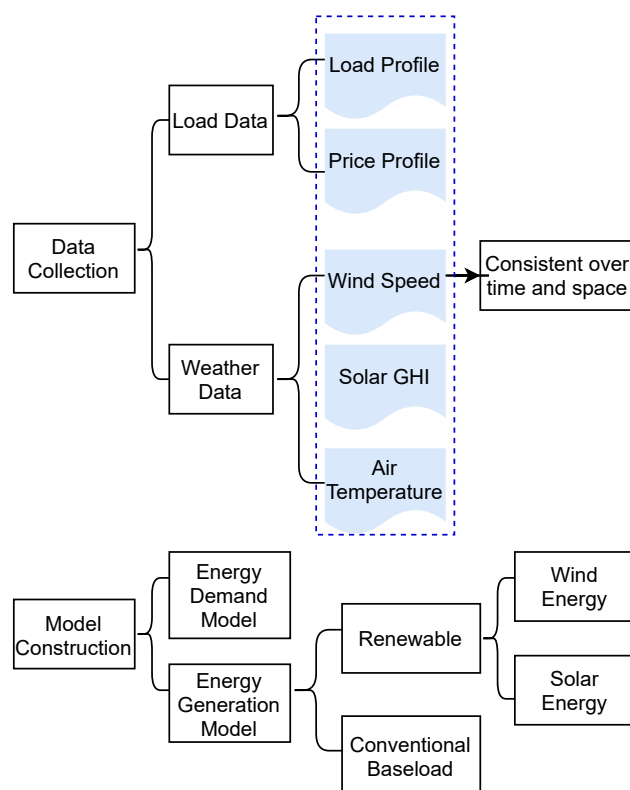


Figure 1. Calculation flow.



**Figure 2.** Step 1 is data collection and model construction.

### 3. Model Construction

This section discusses two models—the energy demand model and the energy generation model. These models represent the basis for generating synthetic time histories for HERON economic evaluation.

#### 3.1. Energy Demand Model

The electricity load data are collected from the Electric Reliability Council of Texas (ERCOT). Historically, from 2010 to 2019, summer peak demand has risen at a 1.4% average annual growth rate (AAGR), and total energy for each year has increased by 2.1%. Reference [18] indicates that the peak demand will be rising at 1.6%, and annual energy will be increased by 2.3% from 2020 to 2029. There are six main factors of forecast uncertainty: weather, economics, energy efficiency, demand response, on-site distributed generation and electric vehicles.

Total load data from the year 2007 to 2013 for seven years from 13 weather zones of Texas are collected as a training set. The price of the electricity data is also collected as an optional correlation variable for the load. The training set is used for feature extraction to generate the ROMs. Several features are extracted, such as the mean and standard deviation of the demand, the Fourier parameters, and so forth. Details on the synthetic time history generation algorithms are discussed in Section 5.

For stochastic optimization calculations, the 1-year ROM is used to generate the synthetic load samples for 120 years, assumed to represent the projected time horizon for the optimization calculations. The samples represent 120 years of operation, with each year emulating the behavior of a single year as obtained from the historical data. HERON allows for capacity expansion over time (i.e., to accommodate projected yearly energy demand increase)s. For Net Present Value (NPV) calculations, it is common practice to perform the initial scoping analysis with no expansion. The projected period of 120 years allows one to take into account for the impact on the time value of money and the depreciation costs in the NPV calculations.

### 3.2. Energy Generation Model

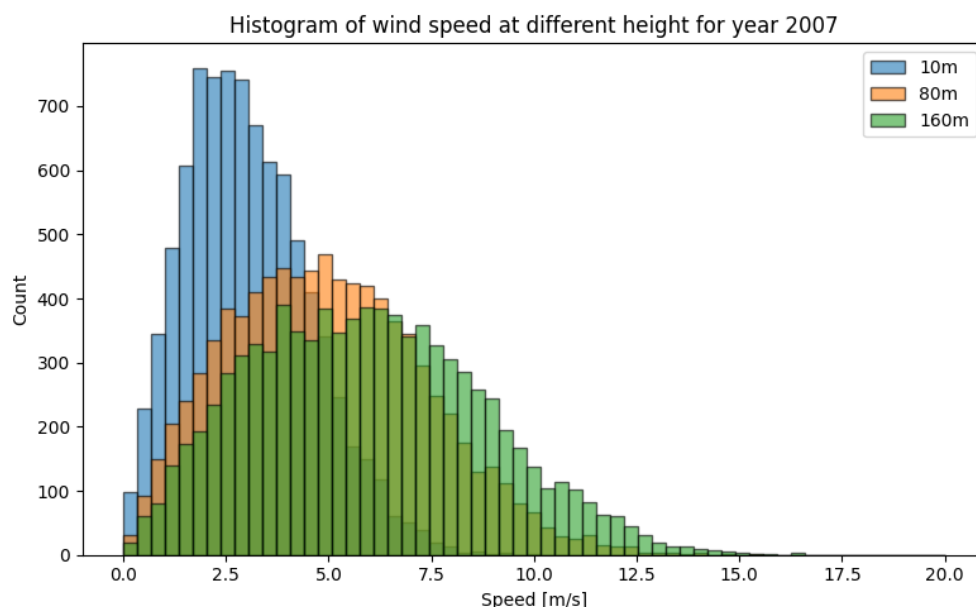
This section discusses the various models employed as a basis for the synthetic histories for the different types of energy units.

#### 3.2.1. Wind Energy Generation Model

The wind energy generation model uses the wind speed and the wind capacity to calculate the wind energy generation. According to a study on wind generation forecasts, core relevant variables are wind speed and direction [19]. The measurement taken at heights nearest to the wind turbines is significantly more relevant than those at other heights.

The Highly Scalable Data Service (HSDS) of the National Renewable Energy Laboratory (NREL) provides wind speed in the Wind Integration National Dataset (WIND) Toolkit [20]. It is the largest publicly accessible meteorological dataset for grid integration. Wind speed at the wanted location has measurements at different heights, including 10 m up to 200 m. However, it only contains seven years of data from the years 2007 to 2013. The time frame covered by these datasets is reasonably recent. The corresponding historical load profile needs to be in the same selected years, so that the wind power and load profile can represent the same weather trends. Both load profile and weather data are highly affected by local weather conditions. Therefore, it is important to prepare the data in the same spatial and temporal resolutions to ensure the consistency of the raw data.

Figure 3 shows three histograms that represent wind-speed hourly measurements taken at three different heights in 2007 in Houston. The graph shows that measurements taken at 160 m have a wider distribution, occurring at a maximum wind speed of 16 m/s, while the measurements at 80m have a maximum wind speed of 14 m/s.



**Figure 3.** Histogram of wind speed at different heights for year 2007.

In support of synthetic time histories generation, a power curve model is used to correlate the wind speed to the generated energy. The wind power curve is adapted from [21]. It is a cubic power model with turbine height at 80 m:

$$P_{wind} = \begin{cases} 0 & U \leq 2 \text{ m/s or } U \geq 18 \text{ m/s} \\ c \cdot U^3 & 2 \text{ m/s} < U \leq 8 \text{ m/s} \\ Pr & 8 \text{ m/s} \leq U < 18 \text{ m/s} \end{cases} \quad (1)$$

$$c = \frac{1}{2} \eta_{max} \rho \pi R^2, \quad (2)$$



where the power curve coefficient  $c$  is 39.06 kg/m, calculated in Equation (2). The  $\eta_{max}$ ,  $\rho$  and  $R$  are the conversion efficiency (0.5926), density of the air (1.17682 g/m<sup>3</sup>) and the radius of the rotor, respectively. The turbine capacity is  $P_r = 20$  kW, the cut in speed is  $u_c = 2$  m/s, the rated speed is  $u_r = 8$  m/s, and the cut out speed is  $u_s = 18$  m/s.

Figure 4 shows the chosen power curve and the frequency distribution of wind speed for 2007. The figure shows that the turbine runs at its capacity around 11% of the time, representing the area under the tail part of the distribution above a wind speed of 8 m/s. Also, the figure shows that the turbine is inactive at low wind speed around 12% of the time, representing the area under the distribution below a wind speed of 2 m/s. This illustrates the fact that, alongside the complexities of wind forecasting, physical operating characteristics also add another source of uncertainty to wind energy production. Note that the wind speed is not steady over the one-hour period, and that the speed changes rapidly with high frequency. The hourly measurement of the speed is used as the average speed over this period. One could even argue that the hourly wind speed may not be the average for that hour at all; however, these short-term fluctuations will not be considered in the current work. This is because our main focus is on the total cost of a combined IES portfolio rather than on the reliability of energy production.

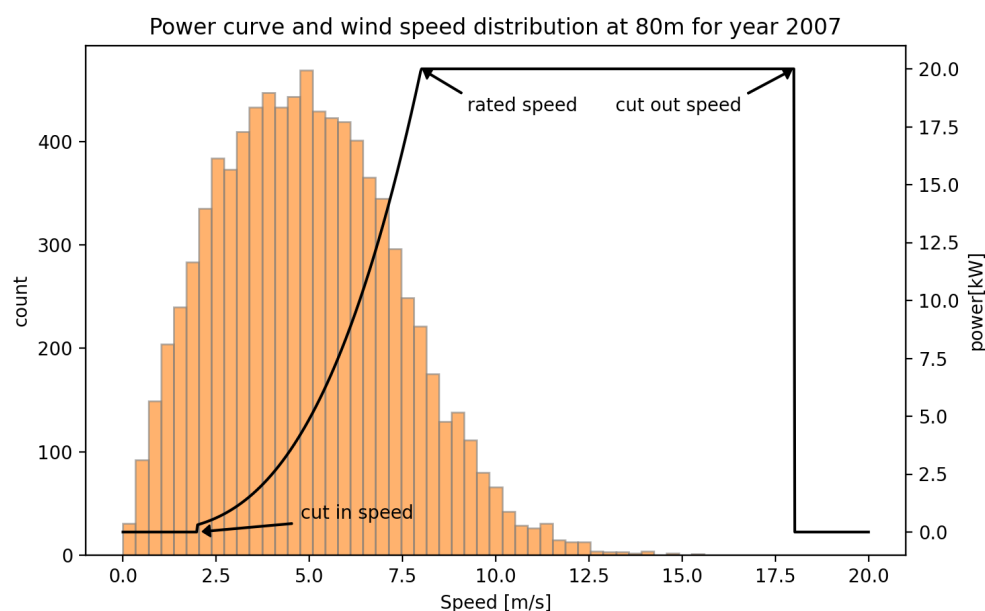


Figure 4. Power curve and the frequency distribution for the year 2007.

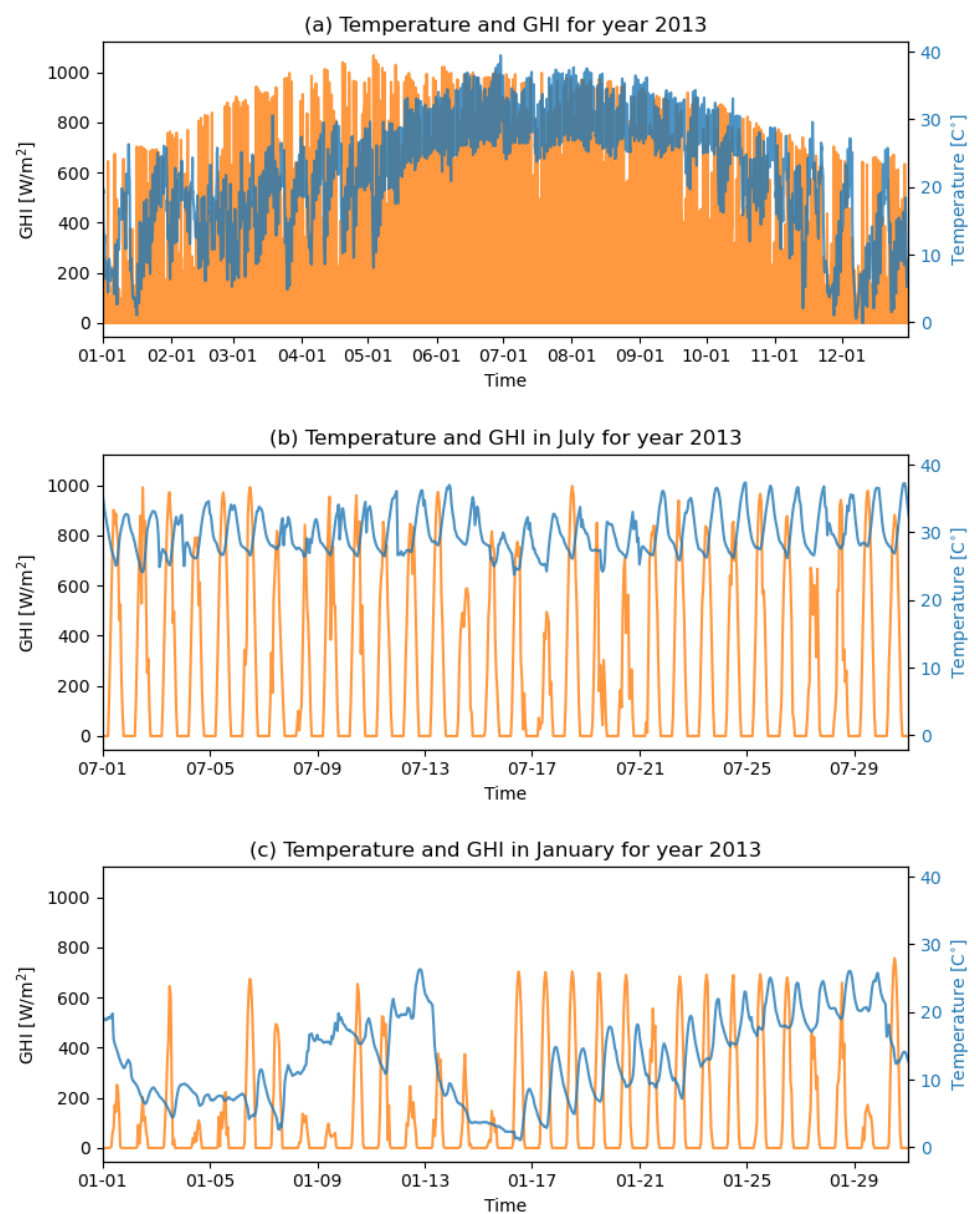
With the wind capacity  $C_{wind}$  fixed,  $\frac{C_{wind}}{P_r}$  is the effective number of turbines, and the corresponding total energy is given by:

$$E_{wind} = P_{wind} \cdot \frac{C_{wind}}{P_r}. \quad (3)$$

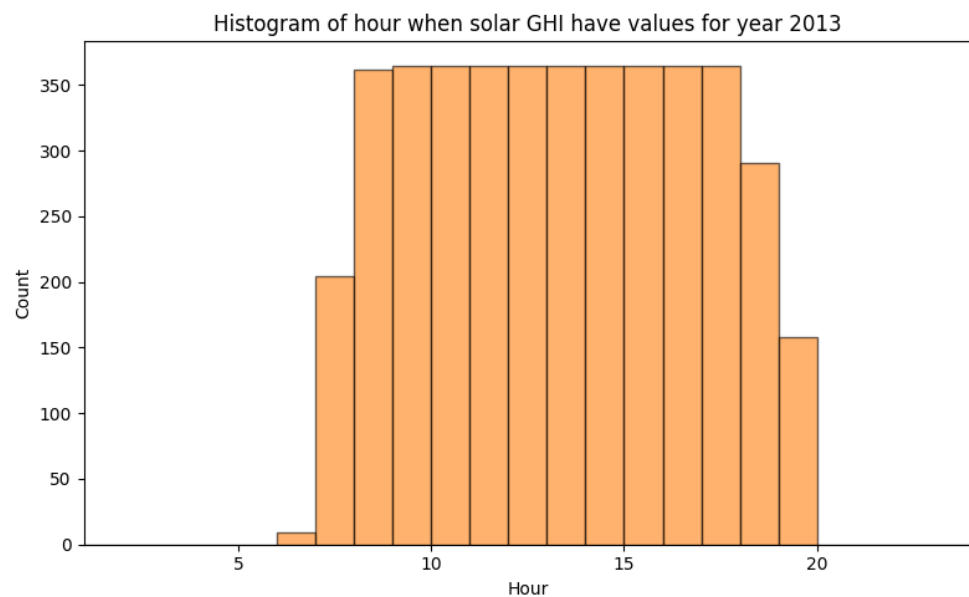
### 3.2.2. Solar Energy Generation Model

Solar photovoltaic devices turn sunlight into electricity. They are employed as the basis for modeling solar energy generation. Solar power performance depends on the incoming radiation and the properties of the solar panel. For productive usage, the maintenance of the energy grid and solar power trading, the prediction of solar energy is very important. Because solar energy generation is related to temperature and solar irradiation, and solar irradiation greatly influences the air temperature, the problem of solar energy prediction is closely related to the problem of weather forecasting. The raw data on Global Horizontal Irradiation and Air Temperature have been obtained from the National Solar Radiation Database (NSRDB) and the WIND toolkit; a scaling of the data was performed to ensure consistency among the collected multi-year data from 2007 to 2013.

Figure 5 shows a merged plot for the air temperature and solar GHI in 2013, followed by two typical zoom-in views over the summer and winter. Analysis of the correlations between these two variables provides insight into the amount of energy generation. For example, in the summer (Figure 5b), when there is a peak in the GHI value, there is a corresponding close-by peak in the air temperature which has a negative impact on the energy generation model (Equation (4)). In the winter, however, this correlation is not as strong, resulting in different amounts of energy generation. Compare the differences between GHI peaks and the air-temperature peak in Figure 5c. Another observation is captured by Figure 6 which logs, in the form of a histogram, the number of hours/day with non-zero GHI values. Results indicate that in 2013, for more than 350 days, the GHI value was non-zero from 9:00 to 18:00, confirming that solar energy provides consistent generation throughout the year.



**Figure 5.** (a) Temperature and solar GHI for year 2013 over the whole year. (b) Zoom-in views over the summer. (c) Zoom-in views over the winter.



**Figure 6.** Histogram of hours when sun is shining from year 2013.

In support of synthetic time history generation, a photovoltaic cell model is employed, which correlates the solar irradiation and the air temperature to generate energy. Solar power generation is adapted from [11,22], with minor improvements from [23]:

$$P_{solar} = \eta \cdot S \cdot \Phi \cdot (1 - 0.005(T - 25)). \quad (4)$$

The  $\eta$  and  $\Phi$  are the conversion coefficient (%) and the solar GHI value ( $\text{kW}/\text{m}^2$ ), respectively. The  $S$  is the exposure area, and  $T$  is the air temperature in Celsius. Notice the negative coefficient for the air temperature which, as discussed earlier, has a negative impact on the energy generation.

With the solar capacity  $C_{solar}$  fixed,  $\frac{C_{solar}}{\max(P_{solar})}$  is the total number of photovoltaic cells, and the total solar energy is calculated by:

$$E_{solar} = P_{solar} \cdot \frac{C_{solar}}{\max(P_{solar})}. \quad (5)$$

### 3.2.3. Conventional Baseload Energies Generation Model

The conventional baseload energy producers considered in our study are natural gas combustion turbines (NGCT), natural gas combined cycle (NGCC), and nuclear energy. They are modeled using two GE LM6000 combustion turbines, model H dual fuel CT in a single-shaft, and two AP1000-type nuclear reactors, as shown in Table 1.

**Table 1.** Plant characteristics.

Energy Unit	Plant Characteristics
Wind	Onshore Wind (WN)
Solar	Utility-Scale Photo-voltaic (PV)
Natural Gas	Combustion Turbine (NGCT) Combined-Cycle (NGCC)
Nuclear	Conventional Nuclear Advanced Nuclear (AN) Small Modular Reactors (SMR)

The NGCT model is adopted from [24] with a nominal output of 100 MW electricity in a simple-cycle configuration. Each turbine is fitted with an evaporative inlet cooler to lower

the inlet air temperature, which is necessary for improving performance in the summer. This NGCT model is based on two aeroderivative dual fuel combustion turbines, each with 53.7 MW power, resulting in a net output of 105.1 MW after deducing the internal auxiliary power.

The NGCC is also adopted from [24]; it includes one combustion turbine, one steam turbine generator and one electric generator. The nominal output for the plant is 430.4 MW.

For nuclear, an advanced nuclear technology is adapted from [25], which is based on the cost estimation of eight companies that have advanced nuclear power plant technology with a capacity greater than 250 MW. Advanced nuclear technologies reflect an evolutionary transition from traditional reactors in terms of safety and non-proliferation, and have a significant role in utility-scale power generation. The cost estimations from some advanced reactor companies all suggest a cost that is lower than the conventional capital cost of nuclear plants.

All conventional baseload producers are assumed to operate at full capacity.

$$E_{\text{baseload}} = E_{\text{capacity}}. \quad (6)$$

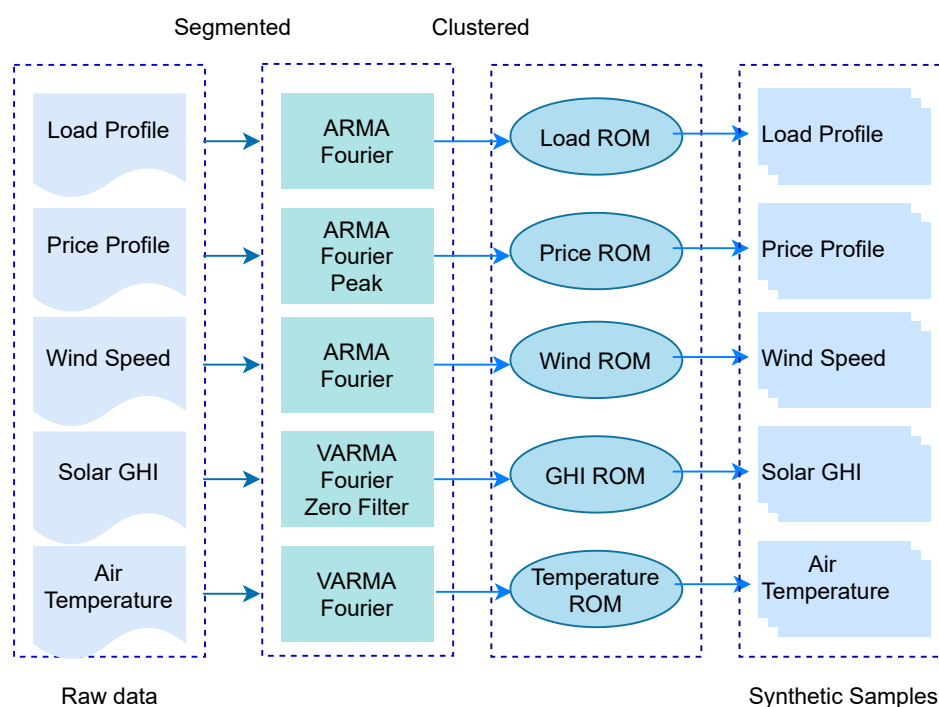
#### 4. HERON Automated Functionalities

This section discusses the second step of the workflow as pointed out in Section 2. Specifically discussed are three key functionalities that are automated by HERON and RAVEN; first, the generation of synthetic data; second, the construction of the energy dispatch model; and last, the cost evaluation of a given mixed-energy production portfolio, respectively discussed in the next three subsections.

##### 4.1. Synthetic Time History Generation

Any techno-economic analysis requires access to representative time history data for the load, demand, and other operational and economic indicators (e.g., pricing data and weather data). If there were infinite records of these historical data, they would be directly used to guide the optimization search. However, in reality, the data are often scarce and only limited to the past few years. The data also exhibit variations on short (i.e., hourly, intermediate, daily and weekly) and longer time scales (i.e., monthly and quarterly), a direct result of the seasonal usage changes. Therefore, it is important to have many representative samples of these time histories to ensure the robustness of the optimization results. To achieve this, the developed workflow relies on the concept of synthetic time history generation. The idea is to construct an ROM, which duplicates the trends (via a process called detrending) and respects the statistical properties identified in the available historical records (via a process called segmentation and clustering). The historical data are, in effect, employed as training data for the ROM model to produce the synthetic time series data. The historical data will be referred to as the training data for the remainder of this report, to distinguish them from the synthetic data generated by the ROM model.

The subsections below provide a brief description of the key ROM algorithms used for generating the synthetic data, as implemented in RAVEN [26,27], also shown as a flow chart in Figure 7. Depending on the type of time series, different ROM algorithms are employed to construct the synthetic time history data (e.g., an ARMA Fourier ROM is used for load profile synthesis, an ARMA Fourier Peak-based model is used for the price profile) as explained in Section 4.1.4. For different historical data, the training process to construct ROM varies. The training process may contain several sub-steps, including segmentation and clustering, Fourier detrending, Auto Regressive Moving Average (ARMA) and peak detection. Fourier detrending is used to capture the seasonal trend, and the ARMA model is employed to describe the stationary residual of the detrended time series.



**Figure 7.** Synthetic time history generation process.

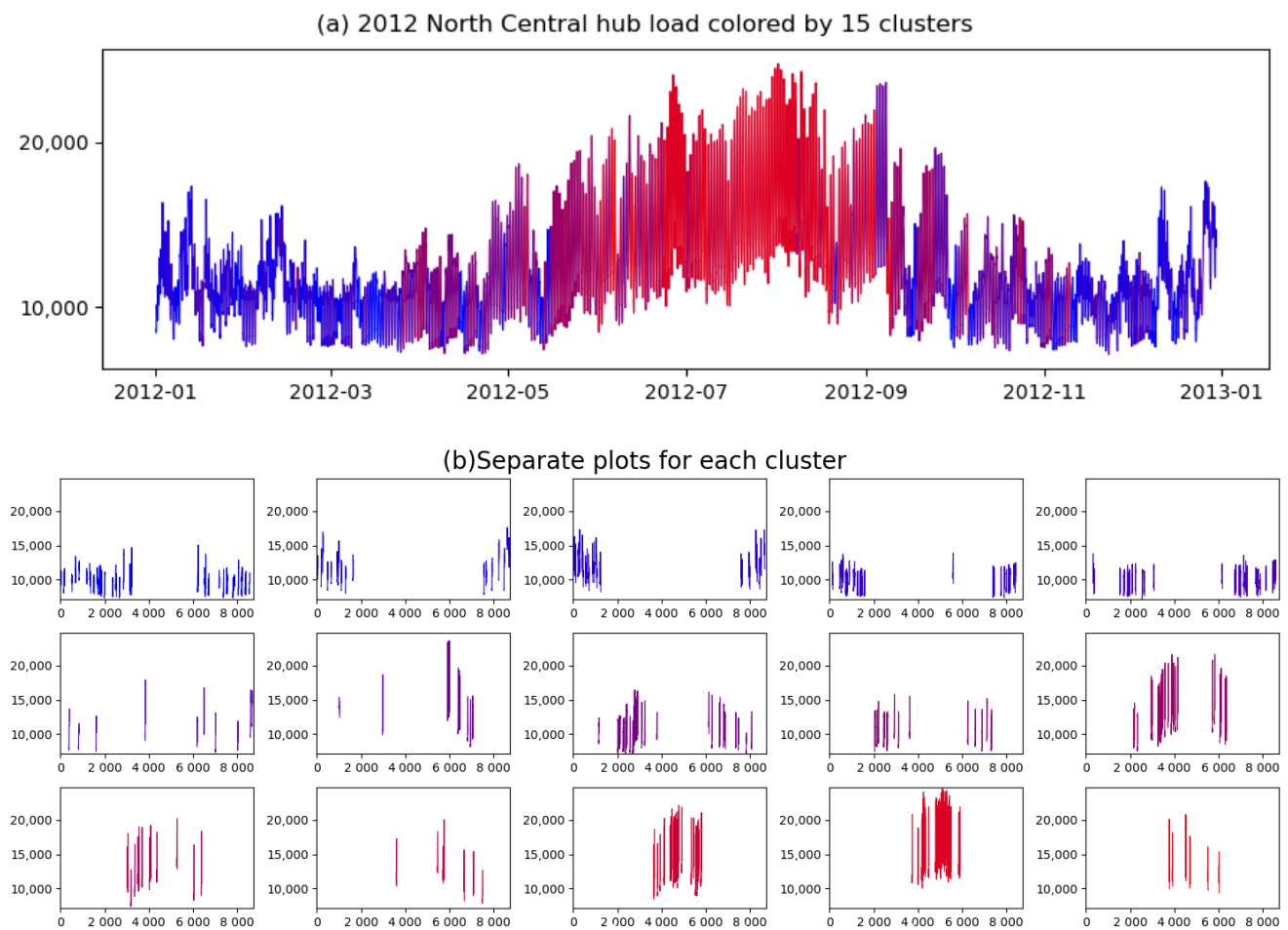
#### 4.1.1. Segmentation and Clustering

Segmentation and clustering are used to define the structure of the time histories in the training data. They can also work as pre-processing steps for other detrending algorithms, and contribute to the detection of significant trends in the training data.

Time series segmentation refers to the process of splitting a time series into segments, defined by  $t_m$ . A time series can be interpreted as a sequence of independent segments of equal length,  $t_m$ , each with its own statistical properties. The objective is to evaluate the time series segment boundaries and describe the complex properties associated with each segment. Different theories exist in the literature regarding the criteria that can be used to decide if the time series can be segmented into regions [28].

Time-series clustering is a common task that seeks to find patterns in the training data to help classify them into distinct groups, based on which synthetic data are generated that respect the statistical features of each group. This work relies on the so-called unsupervised learning algorithm, which does not require labels (i.e., supervision) to identify the best grouping of the data. For more details on the difference between supervised and unsupervised learning, the reader may consult any standard machine-learning textbook [29].

The workflow has tested several potential segmentation and clustering settings, all focused on comparing the performance using different segment lengths (e.g., day, week, month, quarter, and other fractions or multiples thereof). Taking the electricity demand in 2012 in the Texas North Central Hub as an example, the historical load data are shown in Figure 8. A segmentation process employing a one day segment produces 365 segments. These sets are then clustered into 15 smaller sets via a K-means clustering algorithm. A representative result using this segmentation and clustering process is shown in the subplots, with different colors denoting different clusters.



**Figure 8.** (a) 2012 North Central hub load colored by 15 clusters; (b) Separate plots for each cluster.

#### 4.1.2. Fourier

A Fourier detrending algorithm is used to capture the seasonality in the training data. The segmented time histories are decomposed as follows:

$$F_t = \sum_{i=1}^k [a_k \sin(2\pi t_k) + b_k \cos(2\pi t_k)], \quad (7)$$

where  $t_k$  are user-defined time periods and the coefficients  $a_k$  and  $b_k$  are estimated using least-squares linear regression. Note that the time periods  $t_k$  could in general be longer or shorter than the segment's length, defined by  $t_m$ . Next, the fitted Fourier trend is removed from the training time series data, and the residual part is converted into a stationary time series, suitable for ARMA modeling. This is achieved by first converting the residual into a standard normal distribution using a nonlinear transformation, as follows:

$$y_t = \Phi_{normal}^{-1}[f(x_t - F_t)], \quad (8)$$

where  $f$  is a general non-parametric transformation of the residual  $x_t - F_t$ ,  $\Phi$  is a standard normal distribution CDF, and  $y_t$  is the transformed residual time series, to be fitted to an ARMA model.

#### 4.1.3. ARMA

An ARMA model is employed to analyze the transformed time series residuals. This model is used to describe weakly stationary stochastic time series in terms of two polynomials. The first one is the Auto-Regressive (AR) model given as:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t, \quad (9)$$

where  $y$  is obtained from the Fourier detrending and transformation as described above,  $p$  is the number of AR lag terms,  $\phi$  are the ARMA parameters, and  $\varepsilon$  is assumed to be random Gaussian noise.

After adding the Moving Average (MA), the ARMA model can be described as:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (10)$$

where  $q$  is the number of terms in the moving average and  $\theta$  are the weight parameters of the moving average lag term. Next, a least-squares minimization procedure is used to estimate the best values for the  $\phi$  and  $\theta$  parameters.

#### 4.1.4. Peak Detection

The proposed Fourier-based detrending process assumes seasonality that exhibits periodic patterns, with low amplitude and wide peaks. For sharp peaks with high amplitude, a different detrending process is needed, such as the case with daily market price data. To address this need, a new peak detection algorithm is developed to identify and remove the peak, ensuring they are not distorted by the Fourier detrending process. The peak detection algorithm may be abstracted as follows:

1. Let  $x_t$  represent the given training time series data that contains the periodic peak signal. Save the CDF of the training data  $x_t$  as  $\Phi_{x_t}$ .
2. Perform Fourier detrending while limiting the choice of the time periods  $\{t_i\}_{i=1}^k$  to those longer than segmentation length  $t_m$ . This is done to ensure the peak is not distorted by the high-frequency Fourier modes, corresponding to the time periods that are shorter than the segment length. Subtract the fitted Fourier modes to obtain the residual  $\{x_t - F_t^{longer}\}$ , with the superscript denoting that only the longer time periods are used in the detrending process.
3. Divide the residual term  $\{x_t - F_t^{longer}\}$  into  $M$  discrete segments of length  $t_m$ ,  $\{x_i\}_{i=1}^M$ . For each  $x_i$ , collect the peaks' features: peaks' amplitudes, relative location inside the user assigned windows and the probability of each peak's existence, and remove the identified peaks from the residual term. If a peak is found inside a window in the segment, remove the corresponding window from the data.
4. Perform Fourier detrending using the shorter time periods (i.e., those shorter than  $t_m$ ). Save the Fourier coefficients for all the segments.
5. Subtract the fitted Fourier trend from the residual calculated in step 2, to obtain a new residual  $\{x_i - F_{t,i}\}_{i=1}^M$ .
6. Save the CDF of  $\{x_i - F_{t,i}\}_{i=1}^M$  as  $\Phi_{x_i}$  and convert it into a normal distribution (i.e.,  $y_i = \Phi_{normal}^{-1}[\Phi_{x_i}]$ ).
7. Fit the ARMA model for each segment  $\{y_i\}_{i=1}^M$ , save the ARMA parameters for each segment,  $p$ ,  $q$ ,  $\phi_i$  for  $i = 1, \dots, p$  and  $\theta_j$  for  $j = 1, \dots, q$ , to serve as features for the unsupervised clustering algorithm.
8. Generate  $N$  samples of the random Gaussian noise  $(\{\varepsilon_{t,i,j}\}_{i=1}^M)_{j=1}^N$ .
9. Employ the fitted ARMA model to get the  $N$  transformed normal dataset  $(\{y_i\}_{i=1}^M)_{j=1}^N$  and use the inverse distribution function to generate the residuals  $(\{x_i - F_{t,i}\}_{i=1}^M)_{j=1}^N$ .
10. Reconstruct the segments into full-length data, and add the Fourier signal.
11. Add the peaks' signal to the reconstructed data.



All the steps above are automated, except for step 3, which requires a trial and error approach to determine the optimum size window for identifying the peaks. If a small window size is employed, it may not be able to detect the peak and if a wide window is used, the Fourier detrending is expected to distort the shape of the peak, also not allowing its detection.

#### 4.2. Energy Dispatch Model

An energy dispatch model is designed to ensure that the total energy generated by the various types of energy producers meets the demand at the lowest possible cost. This means a strategy that dispatches the maximum amount of energy from the unit with the lowest marginal cost first, before dispatching energy from other units with a higher marginal cost. For example, as shown in the next subsection, nuclear is always dispatched first, followed by NGCC, then NGCT, based on the marginal cost for energy production.

The dispatching decisions are updated every hour and are based on the net load (i.e., the full load minus the load that can be assigned to renewable sources, such as wind and solar, since their marginal cost is assumed to be zero). This assumes that all renewable energy will be dispatched first to the grid before the baseload units. This assumes that there are no penalties for overproduction by renewable sources.

To calculate the net load, the following process is adopted: starting with a synthetic time history for the load, wind speed, solar GHI and air temperature are used to generate synthetic energy generation models for the wind and solar units, which are subtracted from the synthetic load. This results in the net load, given by:

$$\begin{aligned}
 Netload &= Load - E_w - E_s \\
 scale_{Cap} &= \frac{C_n + C_c + C_g}{Max(Net\ load)} \\
 C_{n,c,g}^{new} &= C_{n,c,g}^{old} \cdot scale_{Cap} \\
 E_n &= min_{\sim}(Net\ load, C_n) \\
 E_c &= min_{\sim}(Net\ load - E_n, C_c) \\
 E_g &= min_{\sim}(Net\ load - E_n - E_c, C_g).
 \end{aligned} \tag{11}$$

Subscript  $n, c, g$  means the nuclear, NGCC, and NGCT respectively. The first equation above calculates the net load over the operational horizon, assumed in our model to be 120 years. The scale factor adjusts the initial estimates of the baseload capacities to ensure that the maximum load can be met at any time during the operational horizon. The  $min_{\sim}$  operator is applied on an hourly basis. This implies that, for each hour, the nuclear unit will be dispatched first since it has the lowest marginal cost for electricity generation. If the nuclear unit produces more energy than the net load at any time, the dispatched nuclear energy will be equal to the net load. If the net load is higher than the nuclear capacity, following the same logic, the NGCC unit is dispatched next. If the net load exceeds both the nuclear and NGCC capacities, the NGCT unit is dispatched. It is worth mentioning that dispatch models often allow for some failure probability, that is, the Loss of Load Probability (LOLP); however, this is not explored in the current study.

#### 4.3. Economic Model

This subsection discusses the cash-flow models used to calculate the NPV for the IES model, describing the installed capacities of the various energy producing units. The TEAL plugin (implemented under RAVEN) is employed to automate the cash-flow model calculations for the given IES model. It uses a discounted cash-flow technique [30] to estimate the present worth [31] via:

$$P = \frac{F_t}{(1+r)^t}, \tag{12}$$



where  $P$  denotes the present worth,  $F$  the future worth,  $r$  the annual discount rate, and  $t$  the number of years. Cash flows for the targeted year are called ‘future cash flows’. The NPV is defined as:

$$NPV = \sum_{t=0}^N \frac{CF_t}{(1+r)^t}. \quad (13)$$

The sum runs over the years from 0 to  $N$ . The net cash flows  $CF_t$  are the sum of all cash flows in year  $t$ . The  $N$  is set to 120 years in the current study. Table 2 shows the NPV calculations for each year. It is assumed that the lifetime of the nuclear unit is 60 years, 40 years for NGCT, 30 years for solar and NGCC, and 20 years for wind. The economic model assumes the current grid architecture has no existing generators in place, so for year 0, all plants will be constructed overnight. For the following years, the cash flow will be based on each plant’s fixed operation and maintenance cost (FOM) and variable operation and maintenance cost (VOM). For every ‘building year’ of each energy unit, the cash flow contains the annualized overnight capital cost (Capex) cash flow of the newly built cost and the recurring cash flow for operation and maintenance at the end of the lifetime of this energy unit [16].

**Table 2.** Example cash flows for NPV calculation.

Life Time	Nuclear 60	NGCC 30	NGCT 40	Wind 20	Solar 30
Year					
0	$CF_0^N$	$CF_0^C$	$CF_0^G$	$CF_0^W$	$CF_0^S$
1	$CF_1^N$	$CF_1^C$	$CF_1^G$	$CF_1^W$	$CF_1^S$
...					
20	$CF_{20}^N$	$CF_{20}^C$	$CF_{20}^G$	$CF_{20}^W$ + $CF_0^W$	$CF_{20}^S$
21	$CF_{21}^N$	$CF_{21}^C$	$CF_{21}^G$	$CF_{21}^W$	$CF_{21}^S$
22	$CF_{22}^N$	$CF_{22}^C$	$CF_{22}^G$	$CF_{22}^W$	$CF_{22}^S$
...					
29	$CF_{29}^N$	$CF_{29}^C$	$CF_{29}^G$	$CF_{29}^W$	$CF_{29}^S$
30	$CF_{30}^N$	$CF_{30}^C$	$CF_{30}^G$	$CF_{30}^W$	$CF_{30}^S$ + $CF_0^S$
31	$CF_{31}^N$	$CF_{31}^C$	$CF_{31}^G$	$CF_{31}^W$	$CF_{31}^S$
32	$CF_{32}^N$	$CF_{32}^C$	$CF_{32}^G$	$CF_{32}^W$	$CF_{32}^S$
...					
39	$CF_{39}^N$	$CF_{39}^C$	$CF_{39}^G$	$CF_{39}^W$	$CF_{39}^S$
40	$CF_{40}^N$	$CF_{40}^C$ + $CF_0^C$	$CF_{40}^G$ + $CF_0^G$	$CF_{40}^W$ + $CF_0^W$	$CF_{40}^S$
41	$CF_{41}^N$	$CF_{41}^C$	$CF_{41}^G$	$CF_{41}^W$	$CF_{41}^S$
42	$CF_{42}^N$	$CF_{42}^C$	$CF_{42}^G$	$CF_{42}^W$	$CF_{42}^S$
...					
59	$CF_{59}^N$	$CF_{59}^C$	$CF_{59}^G$	$CF_{59}^W$	$CF_{59}^S$
60	$CF_{60}^N$ + $CF_0^N$	$CF_{60}^C$	$CF_{60}^G$	$CF_{60}^W$	$CF_{60}^S$
...					
119	$CF_{119}^N$	$CF_{119}^C$	$CF_{119}^G$	$CF_{119}^W$	$CF_{119}^S$
120	$CF_{120}^N$	$CF_{120}^C$	$CF_{120}^G$	$CF_{120}^W$	$CF_{120}^S$

With regard to the discount rate  $r$ , energy projects are commonly assessed using discount rates of 3% (cost of capital), 7% (deregulated market rate), and 10% (high-risk investment) [32]. It is known that the social discount rate (SDR) that declines over time is needed. SDRs are used to put a present the value of the costs and benefits that will occur at a later date; a survey of economists indicates that a median SDR of 3% to 7.5% is favorable [33]. Discount rates of 0% to 3% are used in this study for advanced and

conventional nuclear and 3% to 6% discount rates are used for SMR nuclear studies, because our goal is focused on comparing the relative costs of different portfolios with different mixes of renewable and baseload units.

The economic model data sources of the costs are collected from [24,34]. The cost estimates for the FOM and VOM are listed in Table 3.

**Table 3.** Estimates of Cost of Energy Producers.

Unit	Capacity [MW]	Capital [\$ /kW-yr]	FOM [\$ /kW-yr]	Heat Rate [(Btu/kWh)]	Fuel [\$ /MMBtu]	Fuel [\$ /MWh]	VOM [\$ /MWh]	Marginal [\$ /MWh]
NGCC	418	1084	14.10	6431	4.50	28.94	2.55	26.47
NGCT	237	713	7.00	9905	4.50	44.57	4.50	38.58
Wind	200	1265	26.34	/	0	0	0	0
Solar	150	1313	15.25	/	0	0	0	0
Nuclear Conventional	1000	6755	121.64	10,608	0.73	7.61	2.37	9.98
Advanced	250~1000	3782	121.64	/	/	7.00	/	7.00
SMR	300	2600	131.40	/	/	8.00	/	8.00

It is worth mentioning here that existing nuclear power plants are known to have a very high capital cost as compared to NGCC and NGCT plants. Given the renewed interest in nuclear power, studies have been conducted to compare their cost estimates to existing nuclear plants. A survey of these studies indicate that the cost estimate of advanced nuclear plants is almost half that of existing plants. The average capital cost of advanced nuclear [25] is 3782 \$/kW-yr, which is much lower than the corresponding value for an existing plant of 6755 \$/kW-yr. Recently, BWRX-300 SMR was designed by GE, which is estimated to have an overnight cost of less than 3000 \$/kW-yr, for NOAK (nth of a kind) implementations [35]. Another important factor when comparing the various energy units is the cost of fuel, calculated based on the heat rate and thermal energy, as listed in Table 3. For nuclear, fuel price is a relatively small percentage of the overall cost. The VOM cost can include fuel storage, plant decommissioning and waste disposal. The marginal cost is the sum of fuel cost and VOM cost. This study only includes property taxes, which do not change by profit, but only based on the value of the property; it is assumed to be 25.5%. The Modified Accelerated Cost Recovery System (MACRS) is the tax depreciation system in the study [36], which sets the recovery times for depreciation, dependent on life expectancy.

## 5. Optimization Scheme

This section discusses how the values for the installed capacities for the various energy units are optimized to obtain the best NPV for the IES. In principle, one can generate many synthetic samples, as shown earlier for a range of assumed capacities as described by the dispatch model (see Equation (11)), generating a dense cloud of NPVs and picking the set of capacities that provide the best one. This approach, however, is computationally infeasible. Instead, strategies are needed to enable a computationally-efficient search for the optimized capacities. This work proposes an optimization workflow that combines two different methods, the screening curve method and the Gaussian Process regression. The screening curve method provides initial estimates of the optimal capacities assuming a one year operational horizon, and the Gaussian Process model allows one to estimate the NPV for a given set of capacities without redoing the synthetic time histories generation and the TEAL calculations. Each of these two methods is described in a sub-section below. The overall optimization workflow may be described as follows:

1. Generate an  $n$  ordered pair of wind and solar capacities  $C_{renew} = [C_s, C_w]$  using a regular grid structure over a range of their possible/expected values.
2. Use the screening curve method to calculate the optimal baseload capacities  $C_{baseload} = [C_n, C_c, C_g]$  for the given  $n$  samples of wind and solar capacities in step 1.

3. Define the  $i$ th sample  $x_i = [C_n, C_c, C_g, C_s, C_w]$  of the capacities.
4. For each sample  $x_i$ , calculate the NPV cost  $f(x_i)$  by invoking the whole calculation process described before, including synthetic time histories generation, the energy generation model, the energy demand model, the energy dispatch model, and the economic model as automated by HERON and TEAL.
5. Define the matrix of input capacities for all  $n$  samples  $\mathbf{X} = [x_1, x_2, \dots, x_n]^T$ , and a vector of the corresponding NPV values  $f(\mathbf{X}) = [f(x_1), f(x_2), \dots, f(x_n)]^T$ .
6. Train the Gaussian Process model based on the input/output data in step 5.
7. Use the Gaussian Process model to find the capacities corresponding to the best NPV value.
8. To assess the accuracy of the optimal values, generate random  $m$  samples for the capacities, representing random perturbations within the range defined in step 1, denoted by:  $x_j^* = [C_n^*, C_c^*, C_g^*, C_s^*, C_w^*]$ .
9. Calculate the exact NPV values as done in step 4, and find the optimal capacities corresponding to the best NPV value.
10. Compare the optimal capacities from step 7 and step 9.

### 5.1. Screening Curve Method

The screening Curve Method (SCM) is employed to find an estimate of the optimal capacity values, serving as a starting point for the Gaussian Process guided search. The SCM was historically developed to choose an optimal energy portfolio to satisfy the electricity demand [37]. It is based on a single-year operational horizon which limits its value for an IES. For example, SCM cannot optimize the installed capacity of renewable energy because the marginal cost of renewable energy is so low (i.e., renewable energy must be dispatched whenever available). SCM also cannot be easily fitted in multiple energy markets, such as the hydrogen market, which is required in some IESs. Despite these limitations, SCM provides a simple and convenient approach to finding an initial set of estimates for the baseload capacities, which can be used as a starting point for a more elaborate search using the developed Gaussian Process model.

In SCM, the total annual cost for each unit is represented as a function of the firing hour (i.e., the number of hours in which the energy needs to be dispatched by the unit). It combines two curves; the first is called the load duration curve (LDC), representing the dispatched load as a function of the firing hours. It may be thought of as a cumulative density function with the axes reversing their roles. This implies that the y-axis assumes the role of the independent variable for which the PDF is constructed, and the x-axis represents the frequency, that is, the number of hours the load is dispatched. For very high loads, the corresponding frequency is very low, denoting peak times for the load, which do not happen often. However, for very low values of the dispatched load, the frequency is very high, denoting the baseload required throughout the year.

The second curve is the generation cost curve, relating the total annual cost and the firing hour. The y-axis is the total annual cost of the power plant, the intercept of this curve implies the fixed cost of operating the plant and the slope represents the variable cost. This curve determines the total cost if the plant is operated at a fixed dispatched value. The maximum value is reached if the dispatch occurs for all the hours in the year.

A typical SCM curve and its two combined curves are shown in Figure 9. The top red curve represents the nominal LDC curve based on the 2012 historical load data. Given the assumed zero marginal cost for wind and solar, the LDC is adjusted to produce the net LDC curve, which subtracts the load dispatched by wind and solar units. The remaining calculations are based on the net LDC curve, shown in brown.

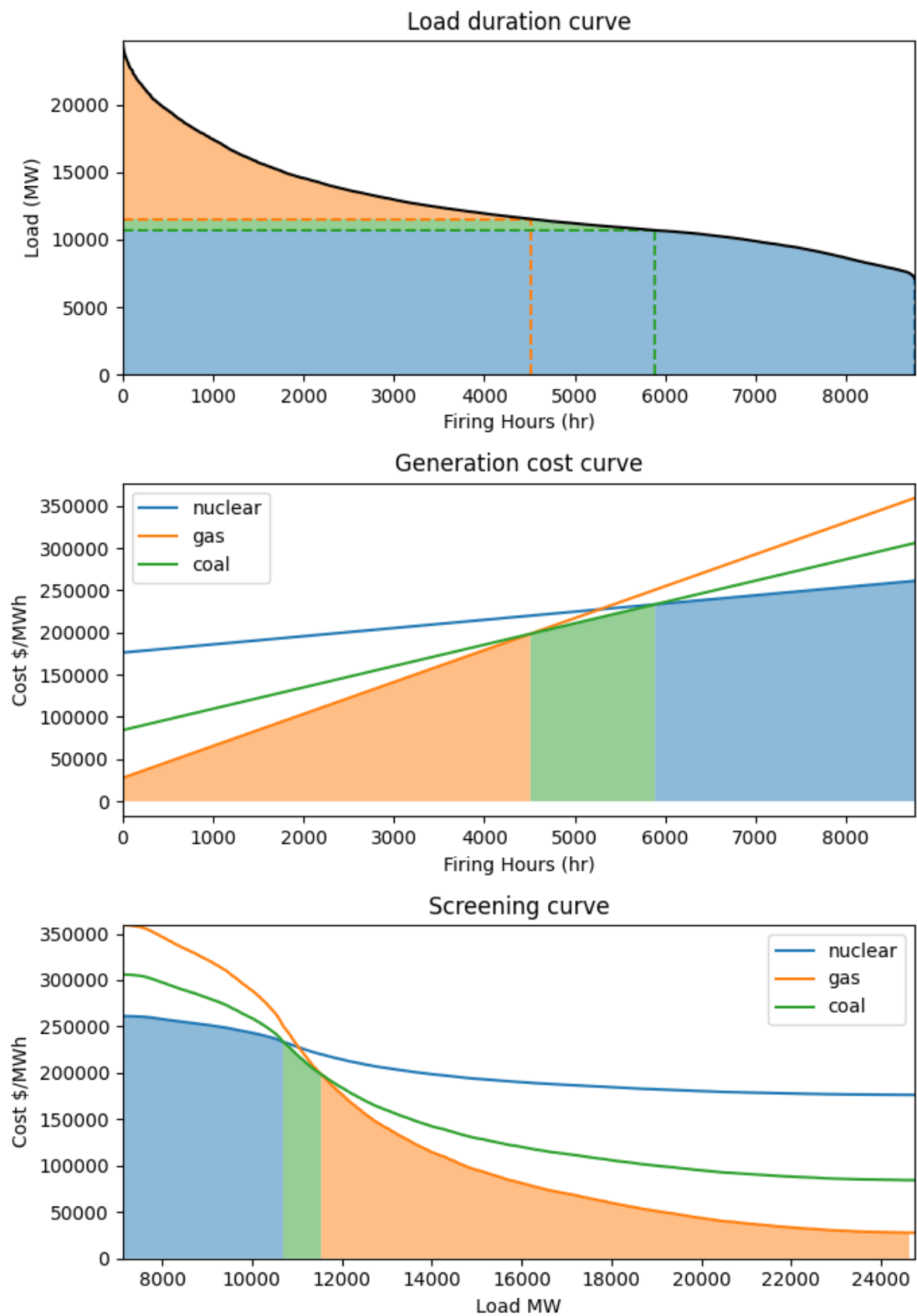


Figure 9. Screening curve for the year 2012 with 8GW solar and 8GW wind capacity with SMR nuclear.

The generation cost curves for the three conventional baseload units are shown in the middle graph. For each unit, the cost includes Capex, FOM per MW capacity per year, VOM cost, and fuel cost (VFOM) per MW of electricity generated per hour. Thus, the total annualized cost of the energy producer can be written as:

$$Cost = Capex + FOM + (VOM + VFOM) \cdot T, \quad (14)$$

where  $T$  counts the firing hours for the given unit. Note that the conventional SCM is based on 1-year data. It is necessary to consider the rebuild and the discount rate. Thus, Equation (14) is replaced by Equation (15). A relatively small discount rate is considered (0~3%) in this study. The annualized capital and operating costs are shown in Table 4.

$$T = Fixed_{Annualized} + Marginal_{Annualized} \cdot T. \quad (15)$$

The middle graph of Figure 9 demonstrates the case without the discount rate, whereby the cost of nuclear is  $177,901 + 8.00 \cdot T$ , the cost of NGCC  $56,173 + 31.49 \cdot T$ , and the cost of NGCT  $27,429 + 49.07 \cdot T$ . The lower envelope curve (tracing the lowest intercept of any vertical line) represents the least-cost solution for a constant number of firing hours. The points on the horizontal axis at which the three curves intersect can be used to determine the best unit for a given number of firing hours. For example, the point of intersection between the cost of NGCT and NGCC is 1635 firing hours, and 5182 firing hours between NGCC and nuclear. This means if the firing hours are less than 4952, the least-cost technology is NGCT, and if the firing hours are between 1635 and 5182, the least-cost technology is NGCC. Nuclear costs the least if the firing hour is more than 5182.

Finally, the bottom graph shows the SCM curve, which is used to determine the optimal mix of capacities considering the variations in the load-firing-hours relationship. Similar to the previous figure, the lower envelope curve determines the best mix of capacities. The first 29.0 GW of the load is best dispatched by the nuclear unit, since they are dispatched for more than 5210 firing hours. The next 12 GW is best dispatched by NGCC, and the last 24 GW is best dispatched by NGCT.

**Table 4.** The annualized capital and operating costs of power plants (SMR nuclear).

Unit	Discount Rate %	Fixed <sub>Annualized</sub> [\$/MW]	Marginal <sub>Annualized</sub> [\$/MW]
Nuclear	0	177,901	8.00
NGCC	0	56,173	31.49
NGCT	0	27,429	49.07
Nuclear	3	60,638	2.16
NGCC	3	23,251	8.50
NGCT	3	13,240	13.24
Nuclear	6	39,844	1.11
NGCC	6	14,964	4.37
NGCT	6	8,976	6.81
Nuclear	9	33,145	0.74
NGCC	9	12,347	2.92
NGCT	9	7,688	4.54

## 5.2. Gaussian Process

Gaussian Process modeling is a well-established area in statistics; it represents a disciplined mathematical approach to build approximations for a process that is statistical in nature. Similar to surrogate modeling techniques, it allows one to train a model based on an available set of input/output data, which can be used later to make predictions. It may be described as a supervised non-parametric regression technique. In our context, this entails an initial training that uses the set of inputs  $[x_1, x_2, \dots, x_n]^T$ , and outputs,  $[f(x_1),$

$f(x_2), \dots, f(x_n)]^T$ , as discussed in Section 5. The goal is to enable one to make predictions for a new value of  $x$  that is not contained in the training set.

A key assumption in Gaussian Process modeling is that the outputs exhibit randomness that is described by a Gaussian distribution with mean values of  $[\mu(x_1), \mu(x_2), \dots, \mu(x_n)]^T$ . Being a non-parametric approach, Gaussian Process relies on a distance metric, which measures the distance between the input samples, assembled in a covariance matrix  $K_{xx}$  of the form:

$$K_{xx} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix} \quad (16)$$

For a new point  $x^*$ , the prediction of  $f(x^*)$  is jointly correlated with the training samples via a joint Gaussian distribution given by:

$$\begin{bmatrix} f(x^*) \\ f(x_1) \\ \dots \\ f(x_n) \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}, \begin{bmatrix} k(x^*, x^*) & k(x^*, \mathbf{x})^T \\ k(x^*, \mathbf{x}) & K_{xx} \end{bmatrix} \right) \quad (17)$$

where

$$k(x^*, \mathbf{x}) = \begin{bmatrix} k(x^*, x_1) \\ k(x^*, x_2) \\ \dots \\ k(x^*, x_n) \end{bmatrix}$$

By marginalizing this joint PDF, the predicted value for the output  $f(x^*)$  is described by a posterior Gaussian PDF:

$$p(f(x^*)|f(\mathbf{x})) \sim N(k(x^*, \mathbf{x})^T K_{xx}^{-1} f(\mathbf{x}), k(x^*, x^*) + k(x^*, \mathbf{x})^T K_{xx}^{-1} k(x^*, \mathbf{x})). \quad (18)$$

The mean  $\mathbb{E}(f(x^*)|f(\mathbf{x}))$  of the posterior can be represented as a linear combination of the kernel function values or the training values:

$$\begin{aligned} \mathbb{E}(f(x^*)|f(\mathbf{x})) &= k(x^*, \mathbf{x})^T K_{xx}^{-1} f(\mathbf{x}) \\ &= \sum_{i=1}^n \alpha_i k(x^*, x_i) \\ &= \sum_{i=1}^n \beta_i f(x_i). \end{aligned} \quad (19)$$

The kernel function is the most important part of the Gaussian process, as it controls the smoothness of the process. It is usually a function of the distance between  $x_i$  and  $x_j$ . Chapter 4 of Ref. [38] provides a detailed example of how to choose the kernel parameter.

## 6. Results

This section demonstrates the applicability of the developed optimization workflow going through all the various steps discussed earlier, including the development of generation models, synthetic time history generation, construction of the SCM curve, and finally the training of the Gaussian Process model.

### 6.1. Synthetic Time History Generation

Figure 10 shows the histograms of the load for all the historical (training) data from 2007 to 2013.

Figure 11 shows the histograms of the synthetic time histories with different training data. Figure 11a collects seven years of synthetic time histories for the load based on the 2012 training data, representing the first seven years of one synthetic sample. The samples

are generated from a trained ROM with each sample synthesizing 120 years' worth of data. Figure 11b shows a similar histogram using 2013 training data. Closer inspection of the synthetic data generated from a single year of training data shows a little volatility from year to year. The distributions, however, are different when the training is based on different years (i.e., 2012 vs. 2013) as shown in the marked differences between Figure 11a,b. For example, comparing to the samples trained from 2013, the samples trained from 2012 have more values from 25 GW to 30 GW, and fewer values from 35 GW to 40 GW. The current study will not explore the impact of these variations on the resulting energy portfolio, but this will be studied in future work.

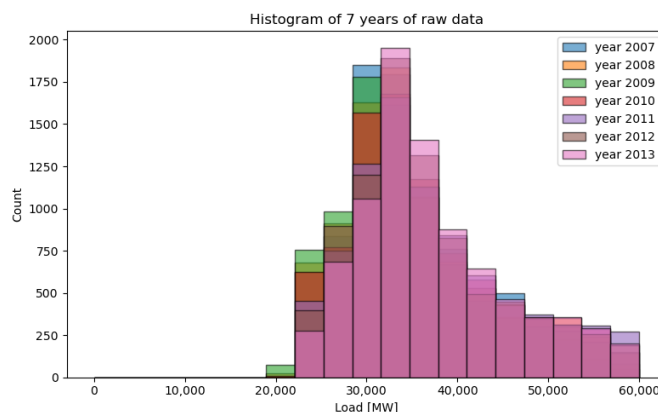


Figure 10. Histogram of hourly load of Texas.

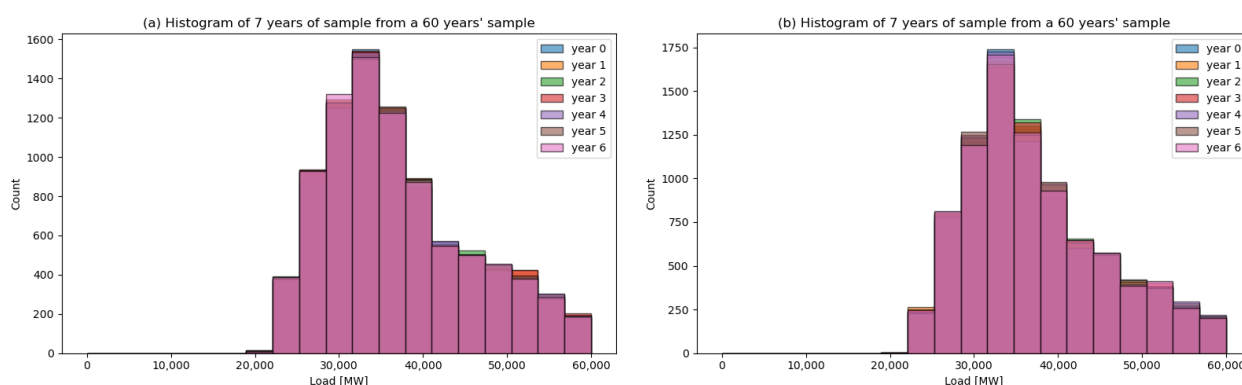


Figure 11. (a) Histogram of synthetic load samples trained from 2012; (b) Histogram of synthetic load samples trained from 2013.

## 6.2. Training Data Results

Figures 12 and 13 show the best cost results for the training data using the SCM with no discount rate with conventional nuclear cost. These results are the initial estimates for the best energy portfolio. Each plot is a heat map on a regular grid of ordered pairs of solar and wind capacities, showing the best NPV results for the combined IES system. The x axis is the solar capacity and the y axis is the wind capacity with a coarse mesh of 500 MW, ranging from 0 to 30 GW.

In 2008, the lowest cost was 12.7 billion dollars, with solar and wind capacities at 10.5 GW and 23 GW. This is the overall lowest cost for all seven years. What stands out in the figures is that the total cost is growing during those years, which is a result of the growing load. It is also apparent from these graphs that, except for an outlier in 2009, the best capacity for solar is in the range of 8 to 12.5 GW, which is around 10% of the overall IES portfolio. However, the best wind capacity ranges from 3.5 GW to 29 GW, which exhibits high volatility. The reason for this is discussed in the following subsection.



The differences between the best wind and solar capacity provide initial estimates about the capacity effectiveness of renewable energy generation.

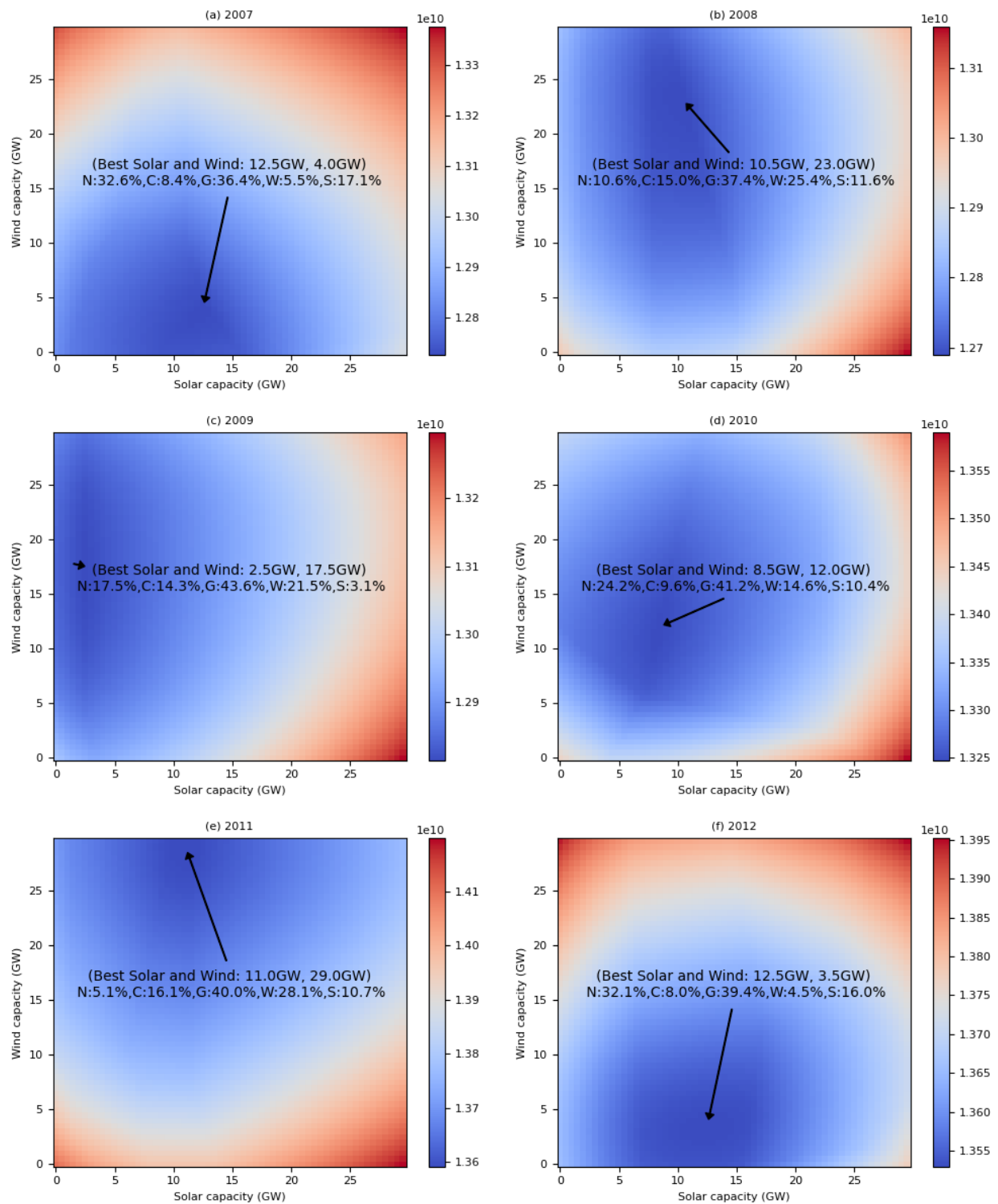
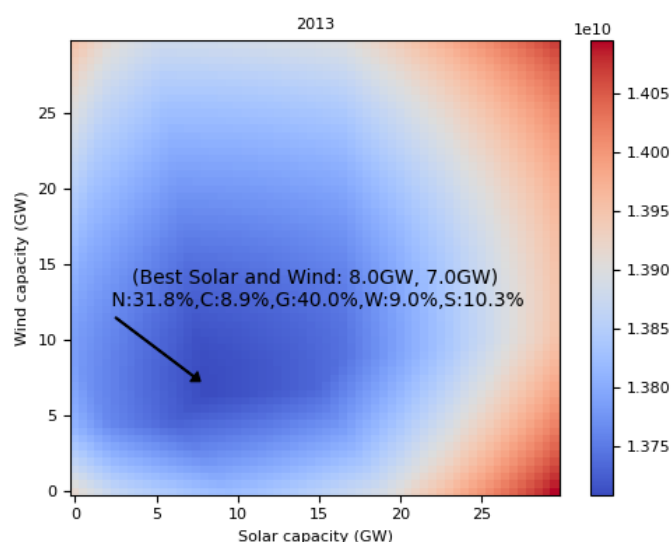


Figure 12. Heat map of cost estimate for (a) 2007 (b) 2008 (c) 2009 (d) 2010 (e) 2011 (f) 2012.





**Figure 13.** Heat map of cost estimate for year 2013.

### 6.3. Effective Renewable Relief for Baseload Generation

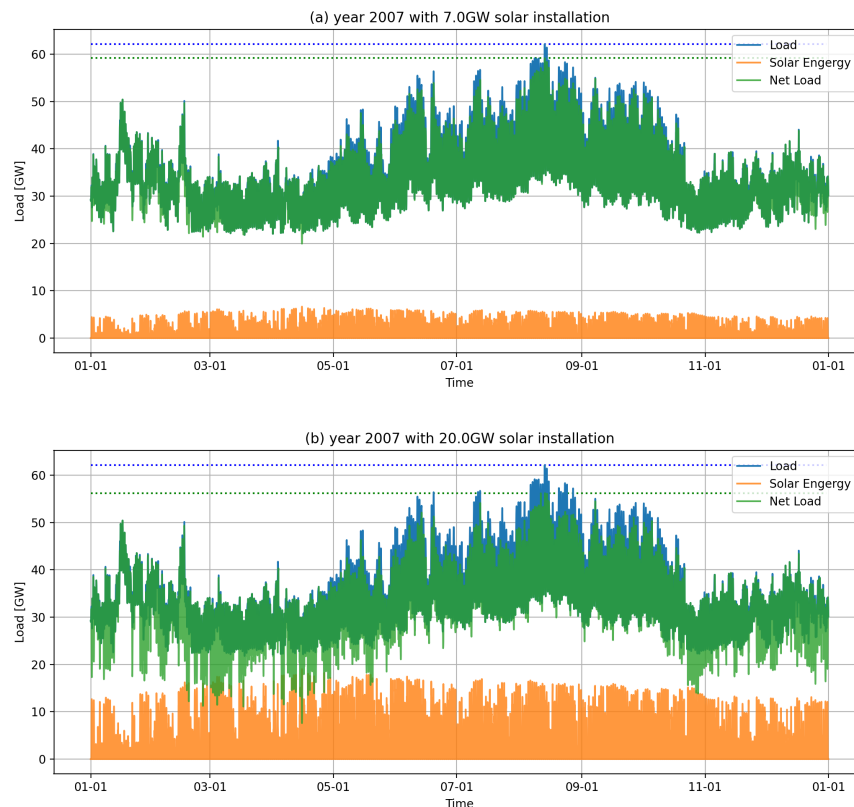
When combining renewable units with baseload units, it is always important to determine whether the increased penetration may have a positive impact on reducing the capital cost for the baseload units (i.e., by providing some relief on their installed capacities). Figure 14 is showing two examples of how much relief on the installed baseload capacities is possible with different capacities of the solar units. The solid blue line shows the original load histories in 2007 as a function of time. The orange line is the solar energy produced, and the green line is then the net load, which is the  $Load - E_{solar}$ .

Figure 14a installs 7GW of solar capacity, and Figure 14b installs 20GW. The dotted horizontal line in each graph is the maximum load and the maximum net load. The difference between the horizontal lines shows the possible reduction in the maximum load to be generated by the baseload units. This reduction (i.e., relief) can be potentially translated into reduced capacities for the baseload units, resulting in capital cost reduction. Recall that in the dispatched model, a scaling factor has been employed to ensure that the maximum load can be met at any time during the operational horizon. So the maximum of the net load determines the total capacity of the baseload units, implying that any reduction in the maximum net load will have a positive economic impact on the IES.

Results indicate that the 7 GW solar installation produces a maximum load relief of approximately 3 GW, while the 20 GW installation provides a relief of 6 GW, implying the law of diminished returns on the investment.

The above results are further detailed in the two subplots of Figure 15, which shows the calculated relief for various combinations of solar and wind capacities. Subplot (a) fixes the wind capacity and varies the solar, and subplot (b) does the opposite. Results indicate that the solar units provide more relief compared to the wind. A 10 GW solar installation can lead to an average of 3 GW effective relief of the load, while wind installation gives an average of 1 GW relief. This is because the energy generation model for the solar unit has a higher correlation with the demand profile, whereas the wind shows more volatility, implying that the peak demand times may not line up with peak production by the wind units, and the production of solar energy provides consistent generation throughout the year.

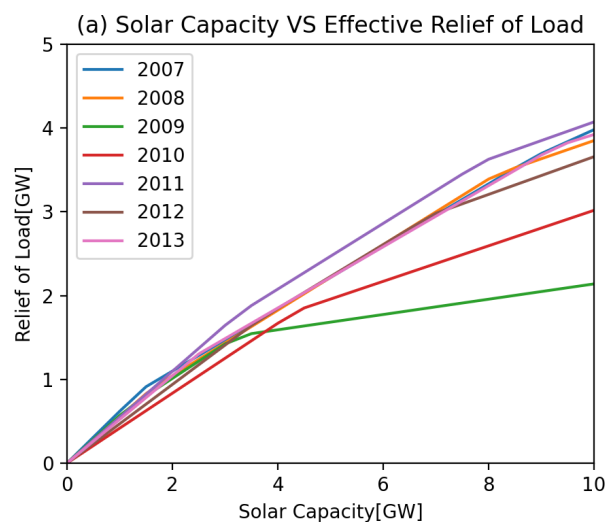
Furthermore, analysis of the subplots in Figure 15 indicate that the initial relief obtained with renewable penetration subsides with their increased capacities. The implication is that wide penetration by renewable is expected to be very taxing in terms of the overall capital cost for the IES.



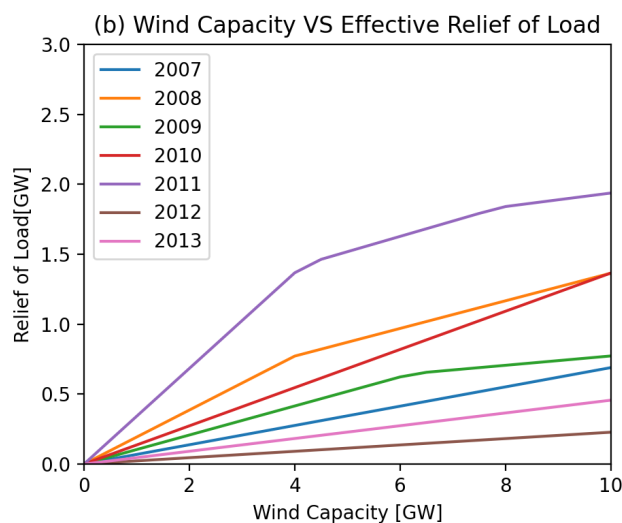
**Figure 14.** Effective renewable relief of installation of solar in 2007 with (a) 7 GW solar installation, and (b) 20 GW solar installation.

In Figure 15a, the green line, which represents 2009, is an outlier from other lines and the growing rate reduces dramatically at around 2 to 3 GW installation. This result matches the observation from Figure 12c, which is the heat map of the least cost using SCM in 2009. Because the effective relief of load for solar is low in 2009, the suggested best capacity for solar is 2.5 GW, this value is relatively low as well.

These trends confirm the optimization results displayed in Figure 12, where it is shown that the wind portion is smaller in 2012 compared to other years, which is confirmed by the low relief obtained for that year as shown in Figure 15b.



**Figure 15.** Cont.



**Figure 15.** Relief of load covered by capital from (a) solar, and (b) wind installation.

#### 6.4. Impact of Economic Model Parameters

This subsection discusses the impact of changing one of the economic parameters—the discount rate—on the optimized capacities. Sample results are shown in Table 5. The goal here is to compare the results for two scenarios, one with conventional nuclear reactors, and one with advanced nuclear reactors.

**Table 5.** Portfolio calculation.

Conventional cost	Nuclear	NGCC	NGCT	Wind	Solar
Gaussian Process	37.6	9.8	45.1	1.5	6.0
$r = 0\%$	32.0	8.7	47.5	2.6	9.2
$r = 1\%$	0.0	40.0	42.7	8.9	8.3
Advanced cost	Nuclear	NGCC	NGCT	Wind	Solar
Gaussian Process	38.5	2.1	45.8	8.5	5.1
$r = 0\%$	38.0	1.7	43.0	9.0	8.3
$r = 1\%$	35.3	2.7	44.1	8.9	8.9
$r = 2\%$	31.7	4.5	46.6	6.4	10.8
$r = 3\%$	20.8	8.8	50.8	1.8	17.7
SMR cost	Nuclear	NGCC	NGCT	Wind	Solar
Gaussian Process	30.4	16.3	29.3	1.9	22.2
$r = 3\%$	34.0	18.0	30.3	3.4	14.4
$r = 6\%$	16.1	31.1	32.6	3.5	16.7
$r = 9\%$	0.0	53.4	37.7	5.0	3.9

If the conventional cost of nuclear 6755 \$/kW-yr is assumed, with a discount rate of 0, the Gaussian Process regression result is consistent with the results of 300 synthetic history samples of 120 years. However, nuclear power's expense grows dramatically as the discount rate rises. The portion of nuclear will be 0 if the discount rate is 1%. This is because the cost of building nuclear overnight is front-loaded and will not be discounted during the 120 year time horizon. However, the rebuild cost for other energy producers will be discounted (see Equation (12)), with the increase in the discount rate ( $r$ ) and rebuild year ( $t$ ), the rebuild cost will decrease exponentially.

If 3782 \$/kW-yr from [25] is used as the cost of nuclear, with a discount rate of 0, RAVEN runs are still consistent with the Gaussian Process results since the changes of the best energy portfolios are within 5%. With the increasing discount rate, the capacity for nuclear capacity reduces, and the solar capacity increases.

If 2600 \$/kW-yr from [35] is used as the cost of nuclear, with the increase of the discount rate from 3% to 9%, wind and gas capacities grow, solar and nuclear capacities reduce. A 9% discount rate result suggests that there should be no nuclear installation.

Based on the December 2020 CDR report [39], wind penetration set a new all-time record for ERCOT, and in 2021 the operational installed capacity in Texas will have 51.0% natural gas, 24.8% wind, 13.4% coal, 4.9% nuclear, 3.8% solar, and 2.1% other energy and storage. There was a significant difference between the 2021 installed capacity and our results. Our study suggests more solar and nuclear capacity, but less wind capacity, because substantial growth in wind capacity might lead to the growth of the total cost or to electricity outages.

## 7. Conclusions

The objective supports one of the key goals for integrated energy systems focused on optimizing the capacities in hybrid energy generation scenarios, and is carried out in a computationally efficient manner. The workflow integrates various key elements to ensure results that are consistent with historical demand data and energy generation, as well as the economic models for the various energy units. Recognizing that a brute force optimization that relies on the analysis of numerous generation scenarios is infeasible, this work builds a workflow that employs a limited set of samples to train a Gaussian Process model, which is more amenable for optimization. The construction of the Gaussian Process model is guided by the Screening Curve Method, a well-proven methodology for portfolio optimization that was developed for the electricity energy market in the 1960s.

The workflow utilized two key plugins in the RAVEN framework, namely HERON and TEAL. HERON automates the energy dispatch calculations based on the given generation model and demand profile, and TEAL is responsible for the economic calculations. Our workflow has employed ROM models to generate synthetic profiles for the load and the renewable energy generation models over a 120 year operational horizon. Different features and detrending algorithms were employed in the construction of the ROM model to ensure all synthetic profiles are consistent with the historical data, including Fourier, ARMA and peak detection-based techniques.

The optimization workflow has been employed to analyze a mixed energy generation portfolio based on the 2007–2013 historical load data in the state of Texas. The IES portfolio includes renewables (e.g., solar and wind units) as well as baseload generators (e.g., nuclear, NGCC, and NGCT units). The results indicate that the solar portion is in the order of 10%, the wind portion shows more volatility from 1% to 10%, nuclear is responsible for approximately one-third of the portfolio, NGCC is in the order of 10%, and NGCT makes up the rest.

Results also indicate that the increased penetration of renewable units is not expected to produce a linear reduction in the IES cost, simply because the solar and wind energy profiles do not correlate well with the demand profile, with solar showing better correlation than wind. The overall implication, however, is that while the increased penetration of renewable sources does indeed reduce the dispatching requirements on the baseload units, it does not reduce the requirements on their capacities, implying that baseload units will have to operate at lower capacity factors; often an undesirable mode of operation for baseload units. Texas experienced blackouts in February 2021, since electricity demands during the extreme cold weather exceeded the prediction, and forced more power sources offline than expected. This is because the natural gas power plants, which generate most of the power in Texas, do not have enough fuel storage on site and rely heavily on the fuel transportation without cold insulation. Although the blackout has been primarily due to issues in the natural gas pipeline system, it is also a result of the fact that the renewable energies, mostly wind generation, were offline.

Finally, future work will focus on developing energy generation models that account for increased energy demand, as well as training using multi-year data, and expand the market analysis to include price and secondary products. Other IES scenarios will also be considered, including energy storage, energy management and process heat applications. Additional studies will also be conducted to study the impact of the clustering parameters

on the quality of the synthesized histories. Some of this work has already been done by the author in support of her PhD [40], but it will be published in a separate article.

**Author Contributions:** Conceptualization, J.Z. and H.A.-K.; methodology, J.Z. and H.A.-K.; software, J.Z., P.T. and C.R.; validation, J.Z.; formal analysis, J.Z. and H.A.-K.; investigation, J.Z.; resources, J.Z., P.T. and C.R.; data curation, J.Z. and C.R.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., H.A.-K., P.T. and C.R.; visualization, J.Z.; supervision, H.A.-K. and C.R.; project administration, H.A.-K. and C.R.; funding acquisition, H.A.-K. and C.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This article was co-authored by employees of Battelle Energy Alliance, LLC (BEA). BEA is the Management and Operating Contractor of the Idaho National Laboratory with the United States Department of Energy (DOE), under Contract Number DE-AC07-05ID14517.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** WIND toolkit Dataset: <https://github.com/NREL/hsds-examples/blob/master/datasets/wtk-us.md> (accessed on 6 April 2021); NSRDB Dataset: <https://github.com/NREL/hsds-examples/blob/master/datasets/NSRDB.md> (accessed on 6 April 2021); Historical electricity load: [http://www.ercot.com/gridinfo/load/load\\_hist](http://www.ercot.com/gridinfo/load/load_hist) (accessed on 6 April 2021); Historical electricity price: <http://www.ercot.com/mktinfo/prices> (accessed on 6 April 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AAGR	average annual growth rate
AN	Advanced Nuclear Facility
ARMA	Auto-Regressive Moving Average
CDF	Cumulative Distribution Function
CF	Cash Flow
CT	Combustion Turbine
ERCOT	Electric Reliability Council of Texas
FOM	Fixed Operation and Maintenance
GHI	Global Horizontal Irradiance
HERON	Holistic Energy Resource Optimization Network
IES	Integrated Energy System
LDC	Load Duration Curve
LOLP	Loss of Load Probability
MACRS	Modified Accelerated Cost Recovery System
NOAK	Nth Of A Kind)
NGCC	Natural Gas Combined Cycle
NGCT	Natural Gas Combustion Turbine
NPV	Net Present Value
NREL	National Renewable Energy Laboratory
NSRDB	National Solar Radiation Database
O&M	Operation and maintenance
PV	Utility-Scale Photo-voltaic Facility
RAVEN	Risk Analysis and Virtual Environment
ROM	Reduced Order Model
SCM	Screening Curve Method
SDR	Social Discount Rate
TEAL	Tool for Economic AnaLysis
VOM	Variable Operation and Maintenance
VRE	Variable Renewable energy
WIND	Wind Integration National Dataset
WN	Onshore Wind

## References

1. Bragg-Sitton, S.M.; Boardman, R.; Rabiti, C.; Suk Kim, J.; McKellar, M.; Sabharwall, P.; Chen, J.; Cetiner, M.S.; Harrison, T.J.; Qualls, A.L. *Nuclear-Renewable Hybrid Energy Systems: 2016 Technology Development Program Plan*; Technical Report; Idaho National Lab. (INL): Idaho Falls, ID, USA; Oak Ridge, TN, USA, 2016.
2. Bragg-Sitton, S.M.; Boardman, R.; Rabiti, C.; O'Brien, J. Reimagining future energy systems: Overview of the US program to maximize energy utilization via integrated nuclear-renewable energy systems. *Int. J. Energy Res.* **2020**, *44*, 8156–8169. [\[CrossRef\]](#)
3. EIA. *Annual Energy Outlook 2019: With Projections to 2050*; Energy Information Administration: Washington, DC, USA, 2019.
4. Koebrich, S.; Chen, E.I.; Bowen, T.; Forrester, S.; Tian, T. *2017 Renewable Energy Data Book: Including Data and Trends for Energy Storage and Electric Vehicles*; Technical Report; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2019.
5. Starn, J. Power Worth Less Than Zero Spreads as Green Energy Floods the Grid. Available online: <https://www.bnnbloomberg.ca/power-worth-less-than-zero-spreads-as-green-energy-floods-the-grid-1.1119243> (accessed on 6 April 2021)
6. Doe, U. *Quadrennial Technology Review 2015*; US Department of Energy: Washington, DC, USA, 2015.
7. Lilly, T. The Big Picture: Nuclear Financial Meltdown. Available online: <https://www.powermag.com/the-big-picture-nuclear-financial-meltdown> (accessed on 1 August 2017).
8. Haratyk, G. Early nuclear retirements in deregulated US markets: Causes, implications and policy options. *Energy Policy* **2017**, *110*, 150–166. [\[CrossRef\]](#)
9. James, R.; Hesler, S.; Bistline, J. *Fossil Fleet Transition with Fuel Changes and Large Scale Variable Renewable Integration*; Technical Report; Electric Power Research Institute: Palo Alto, CA, USA, 2015.
10. Liang, M.; Liu, Y.; Xiao, B.; Yang, S.; Wang, Z.; Han, H. An analytical model for the transverse permeability of gas diffusion layer with electrical double layer effects in proton exchange membrane fuel cells. *Int. J. Hydrogen Energy* **2018**, *43*, 17880–17888. [\[CrossRef\]](#)
11. Tao, C.; Shanxu, D.; Changsong, C. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement. In Proceedings of the 2nd International Symposium on Power Electronics for Distributed Generation Systems, Hefei, China, 16–18 June 2010; pp. 773–777.
12. Epiney, A.; Chen, J.; Rabiti, C. *Status on the Development of a Modeling and Simulation Framework for the Economic Assessment of Nuclear Hybrid Energy*; Technical Report; Idaho National Lab. (INL): Idaho Falls, ID, USA, 2016.
13. Rabiti, C.; Epiney, A.; Talbot, P.; Kim, J.; Bragg-Sitton, S.; Alfonsi, A.; Yigitoglu, A.; Greenwood, S.; Cetiner, S.; Ganda, F.; et al. *Status Report on Modelling and Simulation Capabilities for Nuclear-Renewable Hybrid Energy Systems*; Technical Report; Idaho National Lab. (INL): Idaho Falls, ID, USA; Oak Ridge, TN, USA, 2017.
14. Frick, K.L.; Talbot, P.W.; Wendt, D.S.; Boardman, R.D.; Rabiti, C.; Bragg-Sitton, S.M.; Ruth, M.; Levie, D.; Frew, B.; Elgowainy, A.; et al. *Evaluation of Hydrogen Production Feasibility for a Light Water Reactor in the Midwest*; Technical Report; Idaho National Lab. (INL): Idaho Falls, ID, USA, 2019.
15. Epiney, A.; Rabiti, C.; Talbot, P.W.; Kim, J.S.; Bragg-Sitton, S.M.; Richards, J. *Case Study: Nuclear-Renewable-Water Integration in Arizona*; Technical Report; Idaho National Lab. (INL): Idaho Falls, ID, USA, 2018.
16. Epiney, A.; Rabiti, C.; Talbot, P.; Alfonsi, A. Economic analysis of a nuclear hybrid energy system in a stochastic environment including wind turbines in an electricity grid. *Appl. Energy* **2020**, *260*, 114227. [\[CrossRef\]](#)
17. Talbot, P.W.; McDowell, D.J.; Richards, J.D.; Cogliati, J.J.; Alfonsi, A.; Rabiti, C.; Boardman, R.D.; Bernhoft, S.; la Chesnaye, F.D.; Ela, E.; et al. *Evaluation of Hybrid FPOG Applications in Regulated and Deregulated Markets Using HERON*; Idaho National Lab. (INL): Idaho Falls, ID, USA, 2020. [\[CrossRef\]](#)
18. ERCOT. *2020 ERCOT System Planning Long-Term Hourly Peak Demand and Energy Forecast*; Technical Report; Electric Reliability Council of Texas (ERCOT): Austin, TX, USA, 2019.
19. Castronuovo, E.D.; Usaola, J.; Bessa, R.; Matos, M.; Costa, I.; Bremermann, L.; Lugaro, J.; Kariniotakis, G. An integrated approach for optimal coordination of wind power and hydro pumping storage. *Wind Energy* **2014**, *17*, 829–852. [\[CrossRef\]](#)
20. Draxl, C.; Clifton, A.; Hodge, B.M.; McCaa, J. The wind integration national dataset (wind) toolkit. *Appl. Energy* **2015**, *151*, 355–366. [\[CrossRef\]](#)
21. Lydia, M.; Selvakumar, A.I.; Kumar, S.S.; Kumar, G.E.P. Advanced algorithms for wind turbine power curve modeling. *IEEE Trans. Sustain. Energy* **2013**, *4*, 827–835. [\[CrossRef\]](#)
22. Xiao, W.; Lind, M.G.; Dunford, W.G.; Capel, A. Real-time identification of optimal operating points in photovoltaic power systems. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1017–1026. [\[CrossRef\]](#)
23. Nguyen, D.T.; Le, L.B. Optimal bidding strategy for microgrids considering renewable energy and building thermal dynamics. *IEEE Trans. Smart Grid* **2014**, *5*, 1608–1620. [\[CrossRef\]](#)
24. Cost, C. *Performance Characteristic Estimates for Utility Scale Electric Power Generating Technologies*; US Energy Information Administration, Sargent and Lundy: Washington, DC, USA, 2020.
25. EON. What Will Advanced Nuclear Power Plants Cost? A Standardized Cost Analysis of Advanced Nuclear Technologies in Commercial Development. Available online: <https://www.innovationreform.org/wp-content/uploads/2018/01/Advanced-Nuclear-Reactors-Cost-Study.pdf> (accessed on 20 June 2019).
26. Talbot, P.W.; Rabiti, C.; Alfonsi, A.; Krome, C.; Kunz, M.R.; Epiney, A.; Wang, C.; Mandelli, D. Correlated synthetic time series generation for energy system simulations using Fourier and ARMA signal processing. *Int. J. Energy Res.* **2020**, *44*. [\[CrossRef\]](#)



27. Chen, J.; Rabiti, C. Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems. *Energy* **2017**, *120*, 507–517. [[CrossRef](#)]
28. Keogh, E.; Chu, S.; Hart, D.; Pazzani, M. An online algorithm for segmenting time series. In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 29 November–2 December 2001; pp. 289–296.
29. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*; Elsevier: Amsterdam, The Netherlands; Academic Press: Cambridge, MA, USA, 2009.
30. Higgins, R.C.; Reimers, M. *Analysis for Financial Management*; Numbers 53; Irwin: Chicago, IL, USA, 1995.
31. Dieter, G.E.; Schmidt, L.C. *Engineering Design*; McGraw-Hill Higher Education: Boston, MA, USA, 2009.
32. Varro, L.; Ha, J. *Projected Costs of Generating Electricity*, 2015th ed.; Nuclear Energy Agency (NEA): Paris, France, 2015.
33. Drupp, M.A.; Freeman, M.C.; Groom, B.; Nesje, F. Discounting disentangled. *Am. Econ. J. Econ. Policy* **2018**, *10*, 109–134. [[CrossRef](#)]
34. Irlam, L. *Global Costs of Carbon Capture and Storage*; Global CCS Institute: Melbourne, Australia, 2017.
35. GE. BWRX-300 Overview. Available online: <https://nuclear.gewater.com/build-a-plant/products/nuclear-power-plants-overview/bwr-300> (accessed on 3 February 2020).
36. IRS. *How to Depreciate Property*; Publication 946 (2018); Internal Revenue Service: Washington, DC, USA, 2018.
37. Phillips, D. A mathematical model for determining generation plant mix. In Proceedings of the Third Power Systems Computation Conference, Rome, Italy, 23–27 June 1969.
38. Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
39. ERCOT. *Report on the Capacity, Demand and Reserves (CDR) in the ERCOT Region, 2021–2030*; Technical Report; ERCOT: Austin, TX, USA, 2020.
40. Zhou, J. An Optimization Workflow for Energy Portfolio in IES System. Ph.D. Dissertation, Purdue University, West Lafayette, IN, USA, 2021.