Internship report

Modeling long-term energy storage in the United States

Grid Integration Group (GIG), Lawrence Berkeley National Laboratory

Marc Boëlle, X2021, École polytechnique

Supervisor: Alexandre Moreira, GIG

Contents

Declaration of Academic Integrity

I, Marc Boëlle, hereby declare that this research project report is my own work and that all sources and materials used in its preparation have been acknowledged.

I understand the principles of academic integrity and affirm that this work is free from any form of plagiarism.

Marc Boëlle
July 31st, 2024

I would like to thank Alexandre Moreira, my internship tutor, Miguel Heleno, and the *Grid Integration Group* of the Lawrence Berkeley National Laboratory (LBNL) for their welcome, their help and their invaluable advice in carrying out this research project. I would also like to warmly thank Patricia Silva, PhD student at the Federal University of Juiz de Fora and research affiliate at LBNL, who introduced me to the project.

Abstract

La réalisation de l'objectif des Etats Unis de décarboner à 100 % leur l'électricité d'ici 2030 ([1]) nécessitera de remplacer les combustibles fossiles, tels que le gaz naturel et le charbon, par des technologies non émettrices de gaz à effet de serre pouvant être utilisées à leur plein potentiel grâce aux systèmes de stockage d'énergie classiques et long-terme. Ce rapport, s'appuyant sur les travaux de Patricia de Sousa Oliveira et d'Alexandre Moreira ([2]), étend une étude de cas sur la Californie à tous les Etats contigus des Etats-Unis, en explorant le remplacement d'ici 2050 des générateurs à gaz et à charbon par un mix de sources renouvelables et de systèmes de stockage à long terme. Un coût d'opportunité maximal pour ces technologies est identifié, permettant de les rendre économiquement compétitives par rapport aux projections du mix énergétique de 2050. Une analyse approfondie des résultats est réalisée pour identifier les paramètres influençant les coûts d'opportunité et déterminer les États où l'installation de ces technologies est la plus viable. Ensuite, un nouveau modèle incluant les transmissions d'électricité entre États est proposé pour évaluer leur impact sur les coûts d'opportunité des technologies de stockage longue durée et mieux représenter le marché américain. En raison du coût computationnel élevé lié à la considération simultanée de plusieurs États, un échantillonnage temporel est envisagé pour réduire le nombre de variables et permettre la mise en œuvre prochaine de ce modèle.

Abstract

Achieving the United States' goal of 100% decarbonized electricity by 2030 ([1]) will require replacing fossil fuels, such as natural gas and coal, with non-greenhouse gas emitting technologies that can be used to their full potential thanks to conventional and long-term energy storage systems. Building on the work of Patricia de Sousa Oliveira and Alexandre Moreira ([2]), this report extends a California case study to all contiguous US states, exploring the replacement of gas and coal-fired generators by 2050 with a mix of renewable sources and long-term storage systems. A maximum opportunity cost for these technologies is identified, making them economically competitive with the 2050 energy mix projections. An in-depth analysis of the results is carried out to identify the parameters influencing opportunity costs and determine the states where the installation of these technologies is most viable. Next, a new model including interstate electricity transmissions is proposed to assess their impact on the opportunity costs of long-term storage technologies and better represent the US market. Due to the high computational cost of considering several states simultaneously, time sampling is envisaged to reduce the number of variables and enable the model to be implemented in the near future.

Introduction

Motivation

In 2021, the United States established a long-term strategy to achieve Net Zero Greenhouse Gas Emissions by 2050 [3]. This was all the more reinnforced in May 2023 when the U.S. Department of Energy released a critical report setting a target for 100% decarbonization of the electricity sector by 2030 [1]. This strategy identifies the decarbonization of electricity as the key transformation to reach this goal, focusing on cost-cutting measures for solar and wind power. Focusing on cost-cutting measures for solar and wind power, it also highlights the need to adapt the electricity grid to the flexibility of supply and demand, emphasizing the importance of expanding energy storage systems to address imbalances. Indeed, with the growing share of renewable and inherently intermittent energies, the imbalance between generation and demand is becoming increasingly significant. In regions with high photovoltaic production, energy surpluses during the day can paradoxically lead to electricity curtailments, as these areas may face energy deficits in the evening when solar production decreases.

Energy storage systems are considered critical to achieving this goal. They are used to store surplus electricity during periods of high production and to discharge it during periods of low production, as well as to increase flexibility, as they can provide electricity quickly. These systems have two important characteristics: their power capacity, in MW, which represents the instantaneous power that can be supplied by the system; and their energy rating, in MWh, which denotes the total energy they can discharge before needing to recharge. Duration is defined as the length of time a storage technology can sustain its full rated power output. For long-duration energy storage (LDES), an emerging consensus indicates a minimum duration of ten hours [4]. These systems are being studied because they could discharge energy over an extended period of time when low renewable generation is low, thereby fully compensating for fossil generators removal. the most mature LDES technology at the moment is pumped-hydro storage (e.g. storage in high-altitude lakes), but other technologies are being studied: thermal, electrochemical, mechanical and chemical.

Literature Review and Contribution

Work on defining long-term energy storage is still in progress, as shown in [4]. The authors show the need for 2 types of long-duration energy storage: LDES with a duration of 20 hours to fully correspond to daily cycles, and LDES with a longer duration that would be well-suited for seasonal cycles, by storing energy during periods when production is in surplus (usually summer) and discharging it during the autumn and winter periods, in default. In [5], the authors estimate that with a cost of $20/kW for LDES technology, overall electricity production costs in a power system close to New England could be reduced by 10%. However, like most research to date, this approach relies on technology cost assumptions to identify the impact of LDES on the overall cost of electricity generation. The first study to introduce a framework for determining the opportunity costs of LDES technology was carried out on California by Patricia Silva and Alexandre Moreira([2]) at Berkeley Lab, where I did my internship. These opportunity costs are likely to highly depend on the characteristics of the state. In this report, we present the results of the US national study, which determines the opportunity costs of LDES technology in each state of the contiguous United States. In particular, we analyze how opportunity costs vary according to the state's energy mix, and in which types of state such a transition would be economically feasible. In this work, the states are studied independently, although several studies tend to show that grid transmission can play a major role in the integration of renewables. In [6], the authors identify grid expansion, useful to aggregate generation and distribute resources over large regions, as a key factor. But the role that long-term energy-storage could play in transmission grids is not yet well defined. In this study, we propose a model that could eventually measure the impact of LDES costs within a transmission grid across the United States.

Outline

This report is organized as follows : Section 1 presents the model developed earlier in the team with the modifications I implemented, including the automatic processing of new data to extend the previous study on the state of California to the whole country. In this section, data on generators and storage systems, operational reserves and capacity factor of renewables are derived from several databases presented in 1.4. In Section 2, we present some general considerations on the shape of LDES opportunity costs. The methodology and results of the national study are presented and discussed in Section 3. In Section 4, a refinement of the model is proposed to consider interstate transmissions in order to improve the modeling and study their effect on the costs of long-term storage systems. A time-sampling method is proposed in order to reduce computational requirements, preparing for future results from this model.

Nomenclature

Sets

G Set of indexes of all generators.

G^{cand} Set of indexes of generators that are candidates for investment.

G^{fixed} Set of indexes of fixed existing generators.

G^{firm} Set of indexes of generators able to provide firm dispatchable generation.

G^{renew} Set of indexes of renewable generators.

$G^{firm,fixed}$ Set of indexes of generators equivalent to $G^{firm} \cap G^{fixed}$.

$G^{renew,fixed}$ Set of indexes of generators equivalent to $G^{renew} \cap G^{fixed}$.

$G^{renew,cand}$ Set of indexes of generators equivalent to $G^{renew} \cap G^{cand}$.

$G^{res,providers}$ Set of indexes of generators that can provide reserve.

H Set of indexes of all energy storage systems.

H^{cand} Set of indexes of storage systems that are candidates for investment.

H^{fixed} Set of indexes of fixed existing storage systems.

H^{short} Set of indexes of short-duration storage systems.

$H^{short,fixed}$ Set of indexes of storage systems equivalent to $H^{short} \cap H^{fixed}$.

$H^{long,cand}$ Set of indexes of storage systems equivalent to $H^{long} \cap H^{cand}$.

$H^{short,cand}$ Set of indexes of storage systems equivalent to $H^{short} \cap H^{cand}$.

T Set of indexes of time periods.

Parameters

η_h Round trip efficiency of storage system h.

C^I System power imbalance cost.

C^{short} Reserve shortage cost.

$C_g^{inv,gen}$ Equivalent annual investment cost of candidate generator g.

$C_g^{fom,gen}$ Annual fixed operation and maintenance cost of generator g.

$C_h^{fom,st,power}$ Annual fixed operation and maintenance cost of storage system h.

C_g^p Generation cost of generator g.

$C_h^{st,energy}$ Equivalent annual investment cost in energy capacity for storage system h.

$C_h^{st,power}$ Equivalent annual investment cost in power capacity for storage system h.

C_{gt}^{up} Reserve provision cost of generator g.

$f_{gt}^{available}$ Number between 0 and 1 that determines how much of the generation capacity of renewable unit g is available during time t.

D_t Demand of the system at time period t.

P_g Power generation capacity of generator g.

$P_h^{st,power}$ Maximum power charge/discharge limit for existing storage system h.

$r_g^{up,factor}$ Number between 0 and 1 that determines how much of the generation capacity of unit g can be used for reserves.

$r_t^{up,min}$ Minimum amount of reserve to be held by the system.

RD_g^{factor} Number between 0 and 1 that determines the ramp-down capability of unit g relative to its generation capacity.

RU_g^{factor} Number between 0 and 1 that determines the ramp-up capability of unit g relative to its generation capacity.

S_h Duration of storage system h.

$\underline{V_h}$ Minimum state of charge limit for storage system h.

$\bar{V}_h$ Maximum state of charge limit for storage system h.

$\bar{x}_g^{inv,gen}$ Maximum limit of generation capacity investment for candidate generator g.

$\underline{x}_g^{ret,gen}$ Number between 0 and 1 that determines the minimum reduction in the generation capacity of unit g.

$\bar{x}_g^{ret,gen}$ Number between 0 and 1 that determines the maximum reduction in the generation capacity of unit g.

$\bar{x}_h^{st,energy}$ Maximum limit of energy capacity investment for candidate storage system h.

$\bar{x}_h^{st,power}$ Maximum limit of power capacity investment for candidate storage system h.

$x_h^{st,power\ddagger}$ Predefined power capacity for storage h to be considered in the opportunity value maximization model.

Variables

Δ_t^- Negative power imbalance during time t.

Δ_t^+ Positive power imbalance during time t.

$\delta_t^{up,short}$ Reserve shortage during time t.

c^{BC} Boundary cost of LDES.

p_{gt} Power generation of unit g during time t.

$p_{ht}^{st,ch}$ Power charge of storage h during time t.

$p_{ht}^{st,dis}$ Power discharge of storage h during time t.

$\bar{p}_g^{rem}$ Remaining generation capacity of unit g after reduction.

q^{over} Budget overrun relative to the overall cost determined by the baseline model.

$r_{ht}^{st,up}$ Reserve provisioned by storage h during time t.

r_{gt}^{up} Reserve provisioned by generator g during time t.

v_{ht} State of charge of storage h during time t.

$x_g^{inv,gen\dagger}$ Generation capacity of generator g after investment decision.

$x_g^{ret,gen\dagger}$ Generation capacity of generator g to be reduced after retirement decision.

$x_h^{st,energy\dagger}$ Energy capacity of storage system h after investment decision.

$\bar{x}_h^{st,power\dagger}$ Power capacity of storage system h after investment decision.

Transmission model

N Set of indexes of balancing areas (134 in the contiguous United States).

S Set of indexes of states.

$\mathcal{L}$ Set of indexes of transmission lines between balancing areas.

N_s Set of indexes of balancing areas in state s.

G_n Set of indexes of all generators in balancing area n.

H_n Set of indexes of all storage systems inside balancing area n.

$G_n^{res,providers}$ Set of indexes of generators equivalent to $G_n \cap G^{res,providers}$.

$H_n^{res,providers}$ Set of indexes of storage systems equivalent to $H_n \cap H^{res,providers}$.

$G_n^{firm,fixed}$ Set of indexes of generators equivalent to $G_n \cap G^{firm,fixed}$.

$G_n^{renew,fixed}$ Set of indexes of generators equivalent to $G_n \cap G^{renew,fixed}$.

H_n^{fixed} Set of indexes of generators equivalent to $H_n \cap H^{fixed}$.

$H_n^{long,cand}$ Set of indexes of generators equivalent to $H_n \cap H^{long,cand}$.

G_n^{cand} Set of indexes of generators equivalent to $G_n \cap G^{cand}$.

$H_n^{short,cand}$ Set of indexes of generators equivalent to $H_n \cap H^{short,cand}$.

H_n^{cand} Set of indexes of generators equivalent to $H_n \cap H^{cand}$.

f_l^{max} Maximum power capacity for transmission line $l \in \mathcal{L}$

γ_l Number between 0 and 1 that indicates the efficiency of line $l \in \mathcal{L}$

$f_{n_1 \to n_2,t}$ Power flowing from balancing area n_1 to n_2 during timestep t.

D_{nt} Demand of balancing area n at time t

Δ_{nt}^- Negative power imbalance during time t in balancing area n.

Δ_{nt}^+ Positive power imbalance during time t in balancing area n.

$\delta_{nt}^{up,short}$ Reserve shortage during time t in balancing area n.

c_s^{BC} Boundary cost of LDES in state s.

q_s^{over} Budget overrun relative to the overall cost in state s determined by the baseline model.

τ_t In the time sampled model, duration in hours between time steps t and t+1

1 Review of the opportunity value maximization model

In this section, we recall the methodology of the model used in a previous paper [2] for a case study on California, which we retain for the national case study. We also present the datasets we will be using for this broader analysis.

1.1 Methodology overview

The study focuses on electricity production in the year 2050. It is carried out in two stages:

1. First, an initial model, called **baseline model**, is used to obtain the total annual cost of electricity production without investing in or retiring from any technology;
2. A second model, called **opportunity value maximization model** (OVM), considers the retirement of gas and coal generators, and allows investment in renewable (wind and solar) energies and short- and long-term storage systems. Its objective is to maximize the cost of LDES technology, provided that the total annual cost is lower than that of the baseline.

These two models are presented in more detail in sections 1.2 and 1.3.

1.2 Baseline model

For each state, the baseline model is used to find the overall annual cost of the market clearing problem, i.e. the minimum cost to ensure that generation equals demand at all times. In this model, the costs considered are annual fixed operation & maintenance (FOM) costs, and operation costs as a function of generation at each point in time. We use the data predicted for 2050 by NREL's Cambium described in section 1.4, and so we don't consider any possible investment in new generators. In addition, we consider all generators, including gas- and coal-fired ones. The objective function is as follows:

$$
\begin{aligned}
q^* = \underset{\substack{p_{gt}, p_{ht}^{st,ch}, p_{ht}^{st,dis}, \\ r_{gt}^{up}, \Delta_t^-, \Delta_t^+, \delta_t^{up,short}}}{\text{Minimize}} \quad & \sum_{t \in T} \left[\sum_{g \in G} \left[C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up} \right] + C^{I} \left(\Delta_t^- + \Delta_t^+ \right) + C^{short} \delta_t^{up,short} \right] \\
& + \sum_{g \in G^{firm,fixed}} C_g^{fom,gen} \bar{p}_g^{rem} + \sum_{g \in G^{renew,fixed}} C_g^{fom,gen} \bar{P}_g \\
& + \sum_{h \in H^{fixed}} C_h^{fom,st,power} \bar{P}_h^{st,power}
\end{aligned} \tag{1}
$$

with the global power balance constraint for each timestep :

$$\sum_{g\in G} p_{gt} = D_t + \sum_{h\in H} \left[p_{ht}^{st,ch} - p_{ht}^{st,dis}\right] - \Delta_t^- + \Delta_t^+, \quad \forall t \in T$$
$$\Delta_t^-, \Delta_t^+ \geq 0, \forall t \in T \tag{2}$$

and the reserve requirement constraint for each timestep:

$$\sum_{g\in G^{res,providers}} r_{gt}^{up} + \sum_{h\in H^{res,providers}} r_{ht}^{st,up} \geq r_t^{up,min} - \delta_t^{up,short}; \forall t \in T \tag{3}$$

Other operational constraints used for baseline and OVM models are used in the appendix A.

1.3 Opportunity value maximization model

In this model, we consider the retirement of gas and coal, as well as investment in new technologies. The objective function now reflects the maximization of the cost of the LDES technology, and penalizes exceeding the cost of the baseline model. It is written as:

$$\underset{\substack{c^{BC}, q^{over}, p_{gt}, p_{ht}^{st,ch}, \\ p_{ht}^{st,dis}, r_{gt}^{up}, \Delta_t^-, \Delta_t^+, \\ \delta_t^{up,short}, x_h^{st,energy\dagger}, x_h^{st,power,\dagger}, \bar{p}_g^{rem}, \\ x_g^{inv,gen\dagger}, x_h^{st,power\dagger}}}{\text{Maximize}} \quad c^{BC} - C^{over} q^{over}$$

with, in addition to the operational constraints of the baseline model, the constraint of a lower cost than that of the baseline model. The overall cost of the opportunity value maximization model now takes into account investments in new renewable generators and storage systems, both short and long term. Investment in long-term storage systems is at the c^{BC} dollar cost per kW installed that we seek to maximize. The constraint is as follows:

$$
\begin{aligned}
&\sum_{h \in H^{long,cand}} c^{BC} x_h^{power\ddagger} \\
&+ \sum_{g \in G^{cand}} C_g^{inv,gen} x_g^{inv,gen\dagger} + \sum_{h \in H^{short,cand}} \left[C_h^{st,energy} x_h^{st,energy\dagger} + C_h^{st,power} x_h^{st,power\dagger} \right] \\
&+ \sum_{t \in T} \left[\sum_{g \in G} \left[C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up} \right] + C^{I} \left(\Delta_t^{-} + \Delta_t^{+} \right) + C^{short} \delta_t^{up,short} \right] \\
&+ \sum_{g \in G^{firm,fixed}} C_g^{fom,gen} \bar{p}_g^{rem} + \sum_{g \in G^{renew,fixed}} C_g^{fom,gen} \bar{P}_g \\
&+ \sum_{g \in G^{cand}} C_g^{fom,gen} x_g^{inv,gen\dagger} \\
&+ \sum_{h \in H^{fixed}} C_h^{fom,st,power} \bar{P}_h^{st,power} \\
&+ \sum_{h \in H^{cand}} C_h^{fom,st,power} x_h^{st,power\dagger} \leq q^{*} + q^{over}
\end{aligned}
\tag{4}
$$

where, in [2], the installed capacity $x_h^{power\ddagger}$ is given as input to the model for each long-term candidate. This methodology was chosen to maintain the linearity of the problem. By varying the installed long-term storage capacity and running the model several times, we can obtain a curve of the evolution of the opportunity cost of LDES technology as a function of installed capacity.

1.4 Datasets

1.4.1 ReEDS and PLEXOS models

ReEDS ([7]) is a long-term capacity expansion model for the deployment of electricity generation and transmission systems in the contiguous United States. The model takes into account various factors such as technology costs, policy constraints, and transmission infrastructure. The ReEDS model uses a linear program to make investment and operating decisions to minimize the overall cost of the power system. It considers the contiguous United States, divided into 134 balancing areas linked by transmission lines. In ReEDS, balancing areas are the smallest geographical units where the power balance constraint must be verified. The ReEDS model uses a reduced time resolution of 24 time slices to capture diurnal and seasonal trends.

Once the ReEDS results have been obtained, they are transformed into a database called PLEXOS database. This database is used by the PLEXOS model (see [8]) to create Cambium results.

1.4.2 Cambium database

Based on the results of ReEDS and PLEXOS, the Cambium database ([9]) contains hourly simulation results of energy production over one year, for several scenarios. We use the *Mid_ Case* scenario and the year 2050. It is from this data that we extract temporal data such as demand or renewable capacity factors. It also contains operational parameters such as the capacity of each technology.

1.4.3 Energy prices data

The U.S. Energy Information Administration's Annual Energy Outlook 2023 [10] explores long-term energy trends in the United States, with notable projections influenced by the Inflation Reduction Act. This report includes detailed data on energy and fuel prices by sector and source, which we use for our analysis.

2 General considerations about LDES opportunity cost

First, it is interesting to consider the general shape of the evolution of the boundary cost of the long-term energy storage system as a function of the installed LDES capacity.

Considering a given amount of installed LDES capacity $x_H = \sum_{h\in H^{long,cand}} x_h^{power\ddagger}$, we call

$$z = \begin{pmatrix} (x_g^{\text{inv, gen}\dagger})_{g\in G^{\text{cand}}} \\ (x_h^{\text{st, energy}\dagger}, x_h^{\text{st, power}\dagger})_{h\in H^{\text{short, cand}}} \\ (p_{gt}, r_{gt})_{g\in G, t\in T} \\ (p_{ht}^{st,ch}, p_{ht}^{st,dis})_{h\in H, t\in T} \\ (\Delta_t^+, \Delta_t^-, \delta_t^{\text{up, short}})_{t\in T} \\ (\bar{p}_g^{\text{rem}})_{g\in G^{\text{firm, fixed}}} \\ (x_g^{\text{inv, gen}\dagger})_{g\in G^{\text{cand}}} \end{pmatrix} \in Z_H$$

a state vector where Z_H is the polyhedron of solutions satisfying the operating and investment constraints (identical to those in the baseline). Then, $\forall z \in Z_H$ we call

$$\begin{aligned} C(z) = &\sum_{g\in G^{cand}} C_g^{inv,gen} x_g^{inv,gen\dagger} + \sum_{h\in H^{short,cand}} \left[C_h^{st,energy} x_h^{st,energy\dagger} + C_h^{st,power} x_h^{st,power\dagger} \right] \\ &+ \sum_{t\in T} \left[\sum_{g\in G} \left[C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up} \right] + C^{I} \left(\Delta_t^- + \Delta_t^+ \right) + C^{short} \delta_t^{up,short} \right] \\ &+ \sum_{g\in G^{firm,fixed}} C_g^{fom,gen} \bar{p}_g^{rem} + \sum_{g\in G^{renew,fixed}} C_g^{fom,gen} \bar{P}_g \\ &+ \sum_{g\in G^{cand}} C_g^{fom,gen} x_g^{inv,gen\dagger} \\ &+ \sum_{h\in H^{fixed}} C_h^{fom,st,power} \bar{P}_h^{st,power} \\ &+ \sum_{h\in H^{cand}} C_h^{fom,st,power} x_h^{st,power\dagger} \end{aligned}$$

We're going to prove a few general results about LDES opportunity cost.

Remark In all that follows, we will assume that we choose C^{over} large enough so that if there exists a solution to the optimization problem such that $q^{over} = 0$, then necessarily the result of minimizing $c^{BC} - C^{over} q^{over}$ will be with $q^{over} = 0$. The problem of the opportunity value model is said to be **feasible** if the corresponding cost q^{over} of the solution is zero.

Proposition 1: the problem is feasible if and only if $\exists z \in Z_H, C(z) \leq q^*$, and in this case let's call $z^* \in Z_H$ such that $C(z^*) = \min_{z\in Z_H} C(z)$, we have

$$c^{BC^*} = \frac{q^* - C(z^*)}{x_H}$$

Proof: First, if $\exists z \in Z_H, C(z) \leq q^*$, then $\exists z^* \in Z_H$ such that $C(z^*) = \min_{z \in Z_H} C(z)$ because the optimal cost is upper-bounded by 0 and the problem is linear optimization.

Let's suppose the problem is feasible, which means $q^{over} = 0$. Then the constraint 4 can be written:

$$c^{BC} x_H + C(z) \leq q*$$

For c^{BC} to be maximal, it must be the case that :

1. the corresponding vector z is $z^* \in Z_H$ such that $C(z^*) = \min_{z \in Z_H} C(z)$;
2. the inequality constraint is an equality.

and we obtain the result $c^{BC^*} = \frac{q^* - C(z^*)}{x_H}$.

Reciprocally, suppose $\exists z \in Z_H, C(z) \leq q^*$. Then, taking c^{BC} low enough, the inequality 4 can be verified with $q^{over} = 0$. According to the remark on previous page, the solution to the optimization problem will have $q^{over} = 0$ and so the problem is feasible.

Proposition 2: if there is a capacity $x_{H,1}$ such that the problem is feasible, then for any capacity $x_H \geq x_{H,1}$ the problem is feasible.

Proof: Let x_H be a LDES capacity such that $x_H \geq x_{H,1}$. We note that, since the constraints of the 2 opportunity value model maximization problems are identical except for the maximum amount of investment in LDES, we have $Z_{H,1} \subset Z_H$. Then with $z = z_1^* \in Z_H$, we have $C(z) = C(z_1^*) \leq q^*$. Applying proposition 1, there is a feasible solution with $c^{BC^*} > 0$ and $q^{over} = 0$.

3 National case study

In this section, we first present the methodology used to carry out the study on each state in the contiguous United States. We then present and analyze the results on the opportunity cost of LDES technology when considering a retirement of gas and coal generators. Finally, we discuss the relevance and possible improvements of our model.

3.1 States generation mix presentation

The generators are of three types: gas and coal generators, they are retired in the OVM model; wind and solar generators are renewable generators in which you can invest in the second model to unlock greater power capacity, in addition to the generators you already have; the rest of the generators : nuclear, geothermal, biomass, wind-offshore, hydro generators. They are not retired and their capacity is left identical in the baseline model and the opportunity value model. We can see in figure 1 that the generation mix varies from state to state: for the majority, the mix comes mainly from renewable generators as well as gas and coal, but the split between gas and coal versus renewables varies greatly.

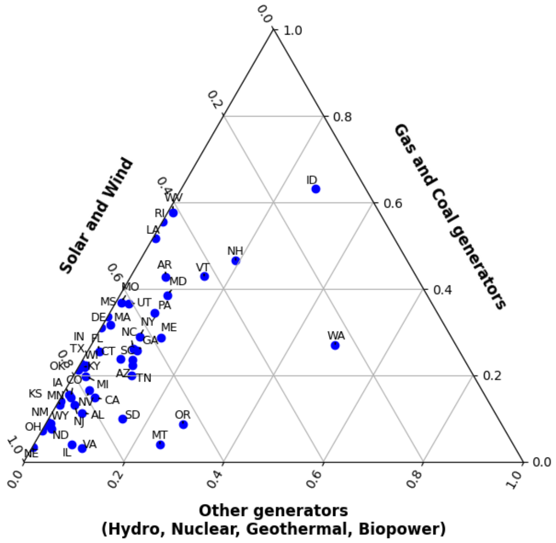


Figure 1: Comparison of the generation mix of states divided into three groups: renewable solar and wind, gas and coal, and others

Similarly, there are 4 types of storage systems:

1. pumped hydro storage (PHS) systems, which store water at altitude and discharge electricity when needed by passing it through turbines;
2. short-duration batteries (2,4,6,8 and 10 hours) that have a round-trip efficiency of 82% in our model.
3. compressed-air energy storage (CAES) that stores energy by compressing air into underground caverns or tanks, which is later released to drive turbines and generate electricity when needed. In Cambium database, there is only one CAES plant, located in Alabama;

4. long-duration energy storage: this technology is only introduced in the opportunity value maximization model. It has a duration of 100 hours, and a fixed capacity for each run. Its round-trip efficiency is 42.5%.

3.2 Methodology framework

The first part of the work involved automating the data pre-processing that had been done manually in the previous study on California, in order to carry out the study on the 48 contiguous US states.

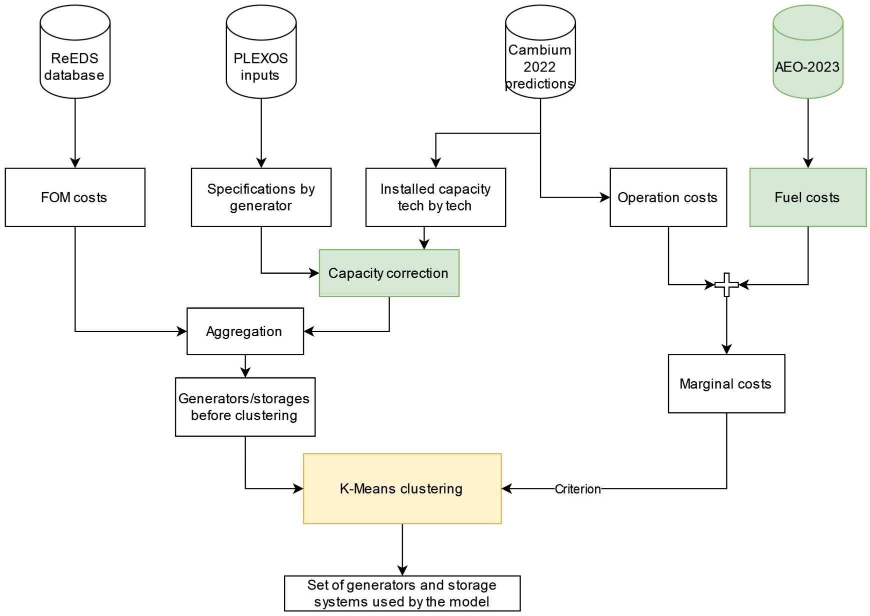
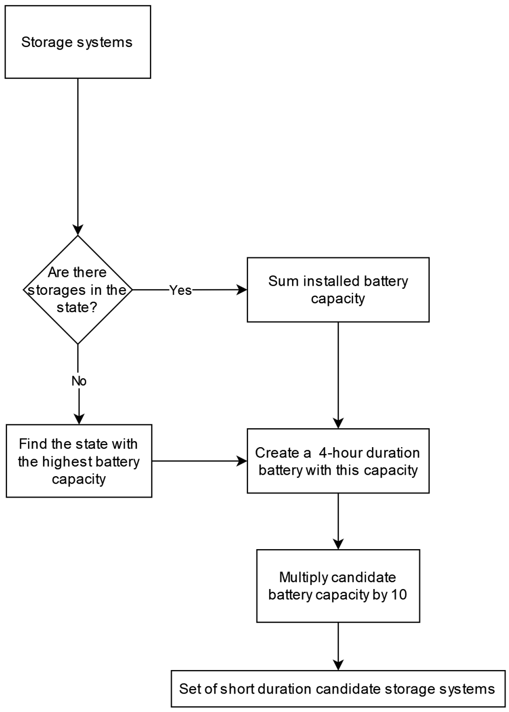


Figure 2: Generators/storage systems processing framework

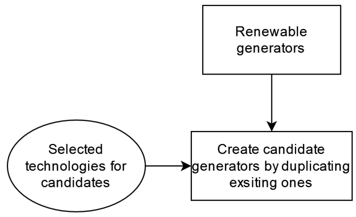
Data pre-processing is shown in figure 2. Generator/storage system data are extracted from the ReEDS database for FOM costs, from the PLEXOS database for individual generator data, and from the Cambium database for production costs and aggregated capacity for each technology. The marginal cost of each generator is calculated by summing the cost of production with the cost of fuel for technologies requiring fuel, reported in \$/kW. To reduce computational cost, k-means clustering is performed on each technology according to variable marginal cost, resulting in a maximum of 3 generators for each technology and balancing area.

Green blocks represent automated elements compared to the previous study. The first one concerns fuel price handling, taken from the AEO 2023 database described in section 1.4.3, obtained by technology and by region of the united states, then extracted to obtain data for each balancing area. The second one concerns capacity correction: in PLEXOS database, we note slightly higher values for the capacity of generators and storage systems compared with the PLEXOS database. In line with what was done in the previous study, we have decided to align ourselves with the capacity of Cambium database's generators/storage systems, applying a correction factor per technology and per balancing area. This was chosen to facilitate comparison between our results and those of Cambium, which are also hourly.

3.2.1 Candidates creation



(a) Candidate storages creation



(b) Candidate generators creation

Figure 3: Methodology for candidates storages and generators in the opportunity value maximization model

In the OVM, investment in new generators and storage systems is possible in place of gas- and coal-fired generators. The methodology for creating candidates is detailed in figure 3. For storage systems, a candidate battery is selected using the following methodology: the candidate battery has a duration of 4 hours and its capacity is the sum of the capacities of all the batteries present in the state. If there are no batteries in the state, a candidate battery is chosen with a capacity corresponding to the sum of the capacities of all the batteries present in the state with the highest capacity for the year 2050, i.e. California. For generators, we have a list of technologies in which it is possible to invest: these are solar and wind (onshore and offshore) technologies . For these technologies, we apply 3b and the corresponding generators are duplicated. As these technologies are already present in every state, this method was always applicable. However, we realized that for some states, the maximum investment in generators was always reached without the annual cost being lower than in the baseline model. For these states, we therefore decided to double the maximum investment capacity in renewables so that the maximum was not reached.

3.2.2 Obtaining target energy demand data

For each state, the target demand is set as the hourly inner generation of the corresponding state found in the Cambium database for year 2050. The inner generation of a state represents the energy production by the generators/storage systems physically in the state. This method allows us to consider a target demand that can be produced even if we restrict the production resources to the state's internal generators and storage systems, and without considering imports or exports from other states. For states bordering Canada, imports from Canada are counted in the inner generation in the Cambium database. Since we are considering the study on a state-by-state basis, and due to the lack of information on Canadian generators in the databases, we have decided to subtract generation from Canadian imports from internal generation in order to determine demand. The resulting electricity demand can be produced by the in-state generators available in the databases used.

3.2.3 Handling missing data

Some generators found in the PLEXOS database are absent from the Cambium database, so some FOM costs are missing. In this case, we choose to replace the missing FOM costs of the generators by those of the same technologies in neighboring balancing areas, on the assumption that these costs are close.

3.2.4 Negative boundary costs

For 7 states, no matter how much LDES capacity is installed, the overall cost is always greater than the baseline cost. These states are Alabama, Connecticut, Delaware, Florida, New Jersey, Ohio and Louisiana. We note that gas- and coal-fired power generation in these states is among the highest in the US, from 20.7% in New Jersey to 77.3% in Delaware.

We then decide to introduce negative opportunity costs to understand how much would have to be invested to install this technology. This cost can be defined as: $c^{BC}x_H = -q^{over}$, or $c^{BC} = -\frac{q^{over}}{x_H}$.

The negative boundary cost curves for each of these states can be found in appendix B.

3.2.5 Optimization environment and tools

To run our models, we used an Intel Xeon CPU E5-2680 v4 2.40GHz, with 64 GB of RAM, using CPLEX 22.1.1 under Pyomo. The algorithms used by CPLEX are primal simplex, dual simplex and barrier algorithm. They are applied in parallel for concurrent optimization, using 1 thread each for primal and dual simplex, and 30 threads for barrier algorithm. The solution is returned by the first algorithm to finish.

As an indication of the size of the programs to solve, for the state of Texas, we have 4,652,060 variables and 8,304,935 constraints for the baseline model, with very little difference for the OVM model. For each state, we run the baseline model once and the OVM model 25 times with different installed LDES capacities.

3.3 Results

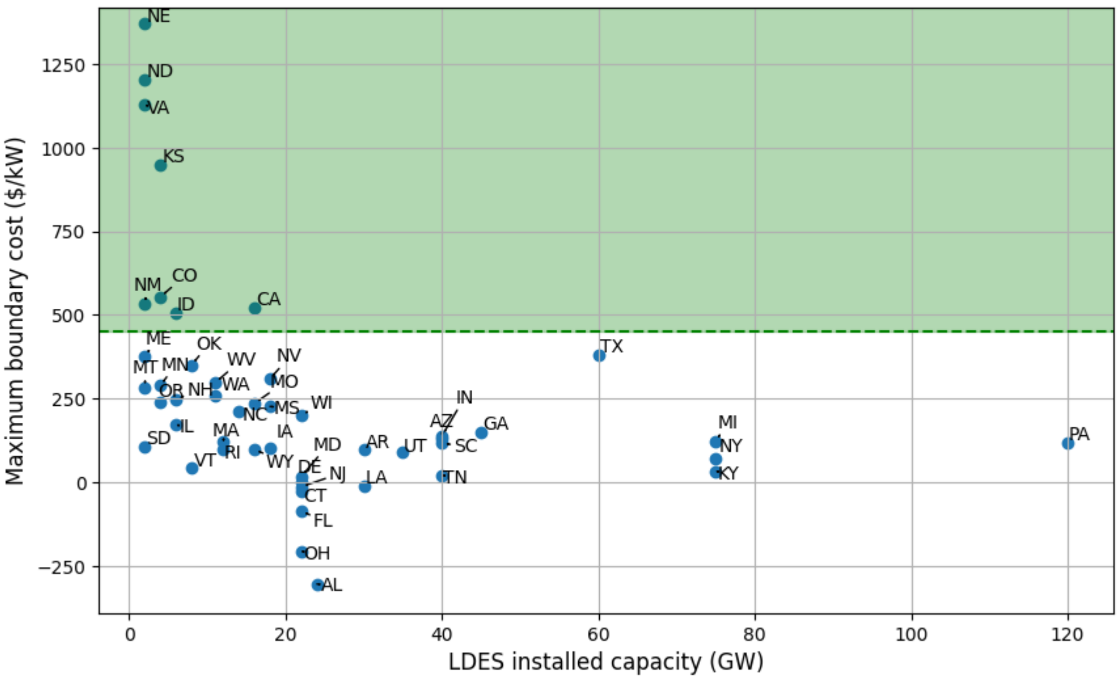


Figure 4: Maximum boundary cost and corresponding LDES installed capacity (MW). The green dashed line of $450/kW comes from the paper [11] and corresponds to the top-quartile forecast of the CAPEX cost (investment cost) of LDES in 2040 (farthest prediction date).

For each state, we display the maximum opportunity cost of LDES technology and the corresponding installed capacity. In figure 4, the green zone corresponds to the area of states with opportunity costs exceeding $450/kW. This prediction by [11] is the top-quartile prediction of investment costs in LDES technology up to 2040. They have been obtained on the assumption that emerging LDES technologies will have similar learning curves to other disruptive energy technologies, such as batteries or offshore wind. According to the trend described in the paper, costs for 2050 are likely to be slightly lower than for 2040. According to figure 4, we distinguish three categories of states:

1. states for which maximum opportunity costs are above $450/kW (green zone), meaning that these states could therefore make a profit by making the transition away from gas/coal generators to more renewables and LDES. The LDES installed capacity corresponding to these high opportunity costs is relatively low compared to other states, ranging from 2.0 GW for New Mexico, Virginia, North Dakota and New England; to 17 GW for California.

2. states whose maximum opportunity cost is in the intermediate zone between $0 and $450/kW.If investment costs remain at or above $450/kW, these states do not have a high enough LDES opportunity cost to make overall costs cheaper than with gas and coal. However, some states (Maine, Oklahoma, Texas) have a maximum opportunity cost not far from $450/kW and could, with adequate subsidy, still make the transition to no gas and coal. It should be noted that the greater the installed capacity (as in

the case of Texas), the greater the subsidies will have to be, since the opportunity cost is per kW installed.

3. states whose opportunity cost could not be calculated, as explained in section 3.2.4, and for which negative opportunity costs are calculated. For these states, there is no optimal boundary cost. We therefore choose the installed capacity of LDES as the average of the capacities of the states for which we obtain the maximum positive boundary cost, i.e. 23.04 GW. For Louisiana, delta minus costs remain at this capacity (see section 3.2.4), so we choose the first capacity at which there are no delta minus costs, i.e. 30 GW.

As LDES can store energy over long periods, we can compare the state of charge of the technology over the months to capture how long-term storage is used in our simulation. In figure 5, we compare in a boxplot, for each state, the monthly state of charge of the LDES (as a percentage of the maximum energy that can be stored) for the LDES capacity such that the opportunity cost of the technology is maximized. First of all, we observe global trends: in the late winter months, particularly February and March, the state of charge is low for most states. In June and July, at the height of summer, the state of charge is often at its highest. This can be explained by the lower availability of renewables generators (especially solar) in winter, so LDES systems have to compensate for this by charging during months of high availability. Moreover, for most states, there is a strong load trend between March and June, while the period between July and January is highly variable depending on the state. Even so, the results are highly variable: not all states store energy with LDES at the same time of year. For this reason, it would be interesting to switch to a case study considering transmissions, since LDES technologies could then also store energy for neighboring regions.

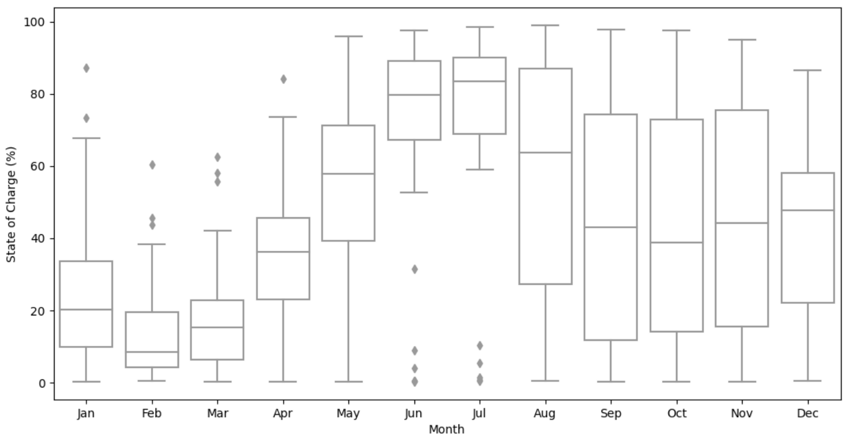
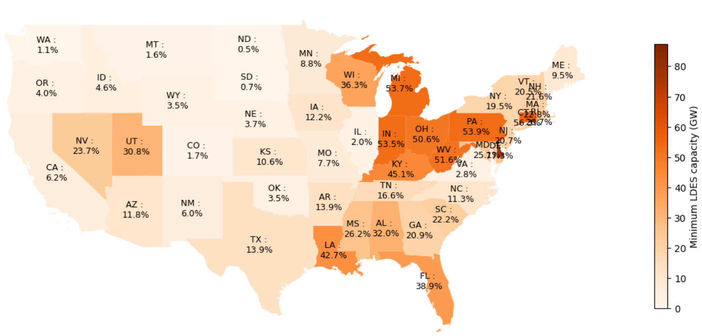


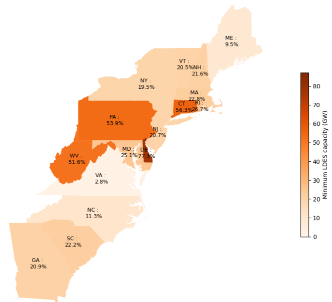
Figure 5: Box plot of the LDES state of charge for each state, for the installed LDES capacity maximizing its opportunity cost

3.3.1 Analysis of the results

We have seen that the opportunity cost results are very different depending on the state. This is not surprising since, as pointed out in figure 1, energy mixes are very different from one state to another. We try to investigate which characteristics have the greatest influence on these opportunity costs. First of all, it should be noted that this cost depends on a very large number of parameters. Each state has its own energy mix, demand, and generators costs. To explain the maximum opportunity cost of the LDES, we choose the following parameters: power capacity for each technology (as a percentage of total power in the state), operating cost of each technology ($/kW), investment cost in candidate technologies ($/MW), average demand over the year and its variance. We also add the power supplied by gas and coal during the year to the baseline model as a percentage of total demand, which can be seen in figure 6 and highlights the disparity between states in terms of their dependence on gas and coal.



(a) Country overview



(b) Zoom on the Eastern coast of the USA

Figure 6: Gas and coal generation in baseline model as a percentage of total demand for each state

To explain our results, we use a regression model. A linear regression to predict opportunity cost is first considered, but given the very low coefficient of determination we try another method: a random forest, which can handle non-linear regression functions. we run the regression on the 41 states with positive LDES opportunity costs. The coefficient of determination obtained is 0.88 .

To determine the importance of the parameters, we use the notion of SHAP value used in game theory ([12]). A SHAP (SHapley Additive exPlanations) value quantifies the contribution of each feature to the prediction of a model, showing the impact of each feature on the model's output. In figure 7, we see these SHAP values for the 5 most explanatory parameters of the model. More precisely, we see the distribution of SHAP values for each feature, where the color and position indicate the magnitude and direction of feature impact across different predictions, with each point representing an opportunity cost prediction's feature contribution. The magnitude of the contribution is directly in the unit of the output variable, here in $/kW.

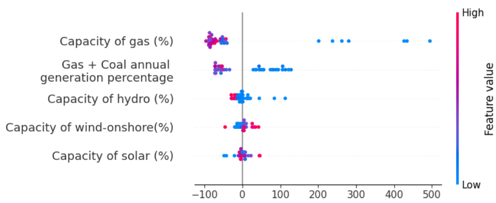


Figure 7: SHAP values of the 5 most explanatory features of the random forest model used to predict the opportunity cost of LDES

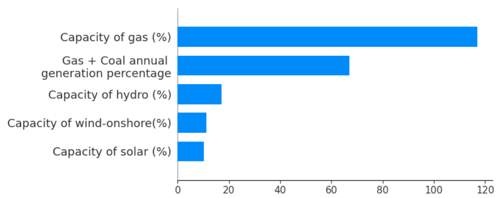


Figure 8: Average of absolute SHAP values (contribution magnitude)

We can see in figure 7 that the lower the gas relative capacity in the baseline model, the higher the corresponding opportunity cost, and conversely. This can be explained as follows: the more gas is used to generate electricity in the baseline, the more investment in renewable generators will be required in the OVM model to compensate for this loss. However, the average cost of investing in wind-onshore is 3.6×10^6 \$/kW, and 1.1×10^7 \$/kW for solar. This compares with the average operating cost of gas 3.2×10^6 \$/kW and 1.1×10^2 \$/kW for coal . Even if solar and wind operating costs are very low, this considerably reduces the margin left for the opportunity cost of LDES. We note that the relative capacity of coal is not among the parameters that most influence the opportunity cost, probably because this technology is present in much smaller capacity than gas in the NREL forecasts for 2050.

We can see in figure 8 that the 3 other variables (relative capacity of hydro, wind-onshore and solar) have a much lower contribution, and it is more difficult to interpret them. For wind-onshore and solar, we can still see that there is a trend towards an increase in opportunity cost as their relative capacity increases (see figure 7), which is consistent since these technologies have a very low operating cost and therefore enable a higher LDES opportunity cost.

Finally, we note that many variables seem to have little impact on opportunity cost, such as demand or the operating costs of different technologies. For demand, it is likely that the opportunity cost of LDES (in \$/kW) will not depend very much on the absolute value of demand, and that it is more the relative characteristics (proportions, percentage) of the energy mix that are important. As for operating costs, they remain fairly similar from one state to another for each technology, and are therefore not decisive. For a possible study in other countries of the world where these costs are likely to differ, it is possible that they play a role in the opportunity cost.

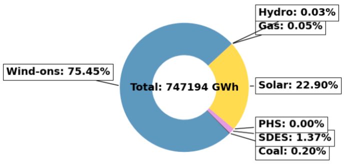
3.4 Discussion and refinements

This study allowed us to estimate the opportunity costs of LDES technology for the 48 contiguous states of the united states. However, there are 2 main limitations: one related to the data and the other to the methodology.

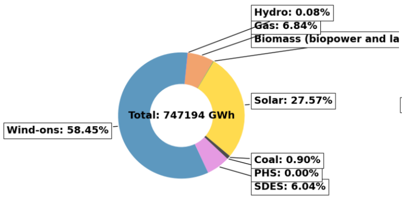
3.4.1 Possible data refinements

For renewable energy capacity factors, we had to slightly modify the methodology used by Patricia Silva in [2] because it was leading to inconsistent results when generalized to other states. The previous methodology consisted in choosing the renewable capacity factors as the maximum between two values : the ReEDS capacity factors, available in the ReEDS database for each region but with a coarse time resolution, and a capacity factor derived from Cambium as follows : for each instant, we compute the ratio between the intermittent generator's production and its maximum capacity, giving a number between 0 and 1. With this methodology, when comparing the baseline model's generation results with Cambium results, we obtained a significant overestimation of capacity for onshore wind power, which completely distorted the energy mix where this technology was widespread (for Texas, in our model, more than 75% of the generation is produced by onshore wind plants in 9a, not even using gas generators which was not consistent with the Cambium results

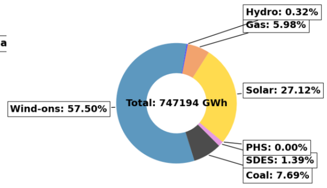
9b). We therefore came back to a methodology only using the capacity factors derived from Cambium database (figure 9c). To justify this choice, we have to assume that each intermittent generator in Cambium database is used to its maximum capacity at each instant. Indeed, if it happens that a generator is not used to its full capacity, then the capacity factor will be underestimated. In practice, the costs of generating intermittent renewable energy are often substantially lower than for other types of generator, which encourages the maximum use of available capacity for these technologies and supports our methodology.



(a) Results of our model using ReEDS and Cambium databases for capacity factors



(b) Results of Cambium database



(c) Results of our model using only the Cambium database

Figure 9: Predicted generation mix for 2050 in Texas as a percentage of total energy production in GWh, with our model (9a and 9c) and with Cambium results(9b)

3.4.2 Absence of uncertainties

In our framework, there is no uncertainty in the model. In power systems, we can identify two types of uncertainties: local uncertainty, i.e. uncertainty about the exact demand within an identified global pattern; and global uncertainty, which affects the choice of global scenario for the study. Examples of such scenarios are “high demand” or “frequent extreme weather events”. We consider only one scenario, which is the MidCase scenario of Cambium 2022. This scenario gives us central estimates for inputs such as technology costs and demand growth. Patricia Silva’s work on the impact on results of scenarios with higher demand, to be published in a future paper, tends to show that the opportunity costs of LDES can be significantly impacted. Secondly, we don’t consider local uncertainties either: the demand and the capacity factors of intermittent generators are predetermined before the optimization. These data are given as inputs to the model, with a demand derived from year 2012 by Cambium (see [9]). This has an impact on the results, as the model optimizes the generation and storage with knowledge of the future.

Inspired by the work carried out in [13] on Southeast Asia, we could take into account both local and global uncertainties. In their study, the researchers associate each global scenario with a probability of occurrence, and then within each scenario, certain parameters such as expected demand can vary within a local uncertainty set. For optimization, the aim is to minimize the cost in the worst-case scenario, i.e. with the local uncertainty set, and then in the scenario leading to the highest cost. With such a method, our model could be more stochastic-robust.

3.4.3 Handling demand and reserve requirements

Our methodology for setting the forecast demand in 2050 for each state consists in setting the target demand equal to the internal generation found in Cambium - allowing to work on each state in isolation. However, it is challenging for 2 reasons. First, this makes it problematic in terms of meeting reserve requirements. According to [14], operating reserves are non-used capacity available for assistance in active power balance. There are many types of operating reserves, the 2 most important being:

- regulating reserves, which are available to correct potential power imbalances of a normal, random nature, and which require an automatic centralized response;
- contingency reserve, a type of reserve that is used in the case of infrequent events creating larger-than-normal variations.

These reserve requirements are expressed as a percentage of demand, and are normally set at around 10-15%, depending on the region. A reserve set at 15% of demand implies that the generators and storage systems could theoretically produce 115% of demand.

However, in the Cambium database, the state's internal generation is based on all internal generators and storage systems available in the state. If at any time, generators are producing at full capacity in Cambium predictions, the internal generation reaches the maximum possible. Then, this target demand in our model becomes too high to simultaneously meet the demand and keep reserve at the same time. This problem arises from the fact that the reserve percentage that generators must maintain is applied to a demand that is not the real one, but a proxy to avoid considering imports and exports and still have a matchable demand. Specifically, in our model, we obtained non-zero Δ_t^- values for several timesteps in the equation 2. This problem occurred for more than 25 states. To avoid it, we adopted the following methodology: for the states and timesteps concerned, we reduced the reserve requirements until all constraints were achievable.

For Virginia, even with 0% reserve requirement, imbalances remained. We therefore reduced demand until the problem could be solved without imbalances, i.e. a 5% reduction in demand.

The second reason why setting the internal generation of each Cambium state as the target demand for our model was challenging, is that when replacing the state's gas and coal generators with a mix of renewable energies and long-term storage systems in the OVM model, we assume that the state's internal generation will remain the same, even though it depends on transmissions between states. But if gas and coal generators are retired in neighboring states, transmissions could be impacted and substantially modify internal generation. The compromise of artificially isolating states energetically creates independence and energy tightness between states to avoid excessive computational requirements. This makes it impossible to identify the effect of implementing long-term energy storage systems throughout the entire territory. This is why the introduction of inter-state transmission to be able to consider imports and exports, and real state demand has been considered in the following section.

4 Towards adding transmission constraints

The conclusions of the previous section showed the limitations of considering states as energetically isolated. In this section, we propose the theoretical foundations of a model that allows modeling with transmissions between mainland states. The computational demands of the model are too great for it to be run directly. We then present a time-sampling solution that reduces the number of variables in the problem.

4.1 Motivations

In the previous model, for each state, the demand is set as the inner generation of a reference energy matrix developed by National Renewable Energy Laboratory's Cambium [9]. This method makes states energy independent of each other, and therefore allows to considerably reduce the number of variables and constraints. However, we obtain results that correspond to an independent energy grid for each state which is not the case in real life. More importantly, it prevents us to consider interactions between states that could impact LDES' boundary costs. For example, the large size of the United States means that sunlight levels, and therefore photovoltaic production, are not identical throughout the country. By transferring energy between zones, we can more accurately represent the energy grid and analyze possible LDES behaviors.

4.2 Modelling

We decided to adapt our case study so as to be able to consider exchanges between regions. We still want to obtain LDES boundary cost results on a state-by-state basis, for comparison with the previous study.

4.2.1 Transmission grid

Like NREL's ReEDS model ([7]), we decide to represent the contiguous United States by a set of 134 zones called balancing areas (BA). These balancing areas are therefore smaller than the states, but respect the state boundaries. The balancing areas and the transmission grid considered by the ReEDS model are shown in figure 10.

For each region and at each timestep, the sum of locally produced power, plus imported energy, minus exported energy, must equal local energy demand, adjusted to take account of energy storage and imbalances. Let N be the set of regions considered for our model. We introduce the set $\mathcal{L} = \{(n_1, n_2) \in N \times N | \text{balancing areas } n_1 \text{ and } n_2 \text{ are electrically linked}\}$. $\forall (n_1, n_2) \in \mathcal{L}, f_{n_1 \to n_2, t}$ represents the energy flow from balancing area n_1 to n_2 during timestep t. The following balance constraint is inspired by equation 2 to take account of these changes:

$$\sum_{g \in G_n} p_{gt} + \sum_{l \in \mathcal{L} | to(l) = n} \gamma_l f_{lt} - \sum_{l \in \mathcal{L} | fr(l) = n} f_{lt} = D_{nt} + \sum_{h \in H_n} \left(p_{ht}^{st,ch} - p_{ht}^{st,dis} \right) - \Delta_{nt}^{-} + \Delta_{nt}^{+}, \forall n \in N, \forall t \in T \tag{5}$$

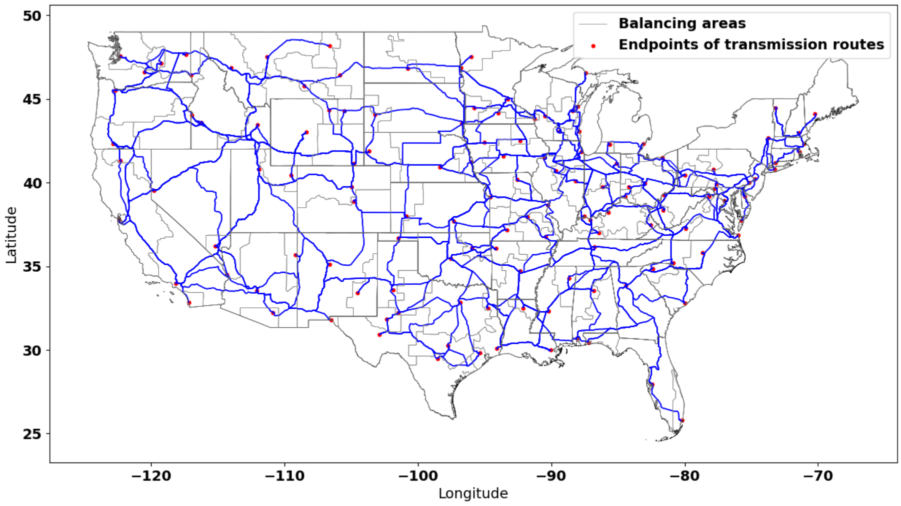


Figure 10: Electrical transmission grid as represented in the ReEDS model on contiguous United States

where γ_l is the efficiency of the line, strictly between 0 and 1. Moreover, energy exchanges between balancing areas are limited by the maximum capacity supported by the grid. For each line, determined by its starting and ending nodes n_1 and n_2, we must have:

$$0 \leq f_{n_1 \to n_2, t} \quad \forall (n_1, n_2) \in \mathcal{L}, \forall t \in T \tag{6a}$$

$$f_{n_1 \to n_2, t} \leq f^{max}_{n_1 \to n_2} \quad \forall (n_1, n_2) \in \mathcal{L}, \forall t \in T \tag{6b}$$

We could add the constraint that $\forall (n_1, n_2) \in \mathcal{L}$, we always have $f_{n_1 \to n_2, t} = 0$ or $f_{n_2 \to n_1, t} = 0$ to force the electric power to flow in one direction only. But with the efficiency $\gamma_l < 1$ in power lines, this is not necessary: Otherwise in the equation 5, it would mean wasting electricity which is never beneficial for minimizing the overall cost of the power system.

For reserve requirements, we adapt the equation 3 on a region-by-region basis by considering that, for a balancing area n, only generators and storage systems within the balancing area can be reserve providers:

$$\sum_{g \in G_n^{res,providers}} r_{gt}^{up} + \sum_{h \in H_n^{res,providers}} r_{ht}^{st,up} \geq r_{nt}^{up,min} - \delta_{nt}^{up,short}; \forall n \in N \tag{7}$$

4.3 Objective function of the baseline and opportunity value maximization models

For the baseline model, optimization is performed once on all states. The only modifications to the baseline model's objective function are to consider power imbalances and reserve shortage for every balancing area:

$$q^* = \underset{\substack{p_{gt}, p_{ht}^{st,ch}, p_{ht}^{st,dis}, \\ r_{gt}^{up}, \Delta_{nt}^{-}, \Delta_{nt}^{+}, \delta_{nt}^{up,short}}}{\text{Minimize}} \quad \sum_{t \in T} \left[\sum_{g \in G} \left[C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up} \right] + \sum_{n \in N} \left(C^{I} (\Delta_{nt}^{-} + \Delta_{nt}^{+}) + C^{short} \delta_{nt}^{up,short} \right) \right]$$
$$+ \sum_{g \in G^{firm,fixed}} C_{g}^{fom,gen} \bar{p}_{g} + \sum_{g \in G^{renew,fixed}} C_{g}^{fom,gen} \bar{P}_{g}$$
$$+ \sum_{h \in H^{fixed}} C_{h}^{fom,st,power} \bar{P}_{h}^{st,power}$$

Since we want an opportunity cost value per state, we will compute the overall annual costs for each state. This will give us a benchmark not to be exceeded in the opportunity value model, where there will be one constraint per state. We note S the set of 48 states of the contiguous USA. Each state s is divided into balancing areas $N_s = \{n \in N | n \text{ is in state } s\}$. After the baseline model has run, we compute the overall costs per state q_s^*, considering the costs of the generators and storage systems in each state:

$$\begin{aligned} q_s^* = \sum_{n \in N_s} \Bigg(& \sum_{t \in T} \left[\sum_{g \in G_n} \left[C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up} \right] + C^{I} (\Delta_{nt}^{-} + \Delta_{nt}^{+}) + C^{short} \delta_{nt}^{up,short} \right] \\ & + \sum_{g \in G_n^{firm,fixed}} C_{g}^{fom,gen} \bar{p}_{g} + \sum_{g \in G_n^{renew,fixed}} C_{g}^{fom,gen} \bar{P}_{g} \\ & + \sum_{h \in H_n^{fixed}} C_{h}^{fom,st,power} \bar{P}_{h}^{st,power} \Bigg) \end{aligned} \tag{8}$$

Then, in the opportunity value maximization model, we will define as many opportunity cost variables c_s^{BC} as states. The new opportunity value maximization objective function is:

$$\underset{\substack{c_s^{BC}, q_s^{over}, f_{lt}, p_{gt}, p_{ht}^{st,ch}, \\ p_{ht}^{st,dis}, r_{gt}^{up}, \Delta_{nt}^{-}, \Delta_{nt}^{+}, \\ \delta_{nt}^{up,short}, x_h^{st,energy\dagger}, x_h^{st,power,\dagger}, \bar{p}_g^{rem}, \\ x_g^{inv,gen\dagger}, x_h^{st,power\dagger}}}{\text{Maximize}} \quad \sum_{s \in S} \left(c_s^{BC} - C^{over} q_s^{over} \right)$$

subject to :

$$
\begin{aligned}
&\sum_{n\in N_s}\Bigg(\sum_{h\in H_n^{long,cand}} c^{BC} x_h^{power\ddagger} \\
&\quad+\sum_{g\in G_n^{cand}} C_g^{inv,gen} x_g^{inv,gen\dagger} + \sum_{h\in H_n^{short,cand}} \left(C_h^{st,energy} x_h^{st,energy\dagger} + C_h^{st,power} x_h^{st,power\dagger}\right) \\
&\quad+\sum_{t\in T}\left[\sum_{g\in G_n}\left(C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up}\right) + C^{I}(\Delta_{nt}^{-} + \Delta_{nt}^{+}) + C^{short}\delta_{nt}^{up,short}\right] \\
&\quad+\sum_{g\in G_n^{firm,fixed}} C_g^{fom,gen}\bar{p}_g^{rem} + \sum_{g\in G_n^{renew,fixed}} C_g^{fom,gen}\bar{P}_g \\
&\quad+\sum_{g\in G_n^{cand}} C_g^{fom,gen} x_g^{inv,gen\dagger} + \sum_{h\in H_n^{fixed}} C_h^{fom,st,power}\bar{P}_h^{st,power} \\
&\quad+\sum_{h\in H_n^{cand}} C_h^{fom,st,power} x_h^{st,power\dagger}\Bigg) \le q_s^{*} + q_s^{over}, \forall s\in S
\end{aligned}
\tag{9}
$$

4.3.1 Methodology for inputs

A region's demand can now be matched not only by internal generators and storage systems, but also by energy imports. This is why it is possible to use the Cambium database busbar load for 2050, which is the total electric load in a region. For balancing areas bordering Canada, we can set demand as the busbar load minus Canadian imports, since we still do not consider Canadian generators and we do not want the demand to be too high to be matched. Sometimes this method gives slightly negative demand because Canadian imports can be transmitted to other boundary areas. In this case, we discard the negative demand and put it to 0.

For power transmission line capacity data, NREL's Standard Scenarios 2023 data ([15]) is easy of use, as the spatial division of the ReEDS model is also used. Using a capacity expansion model, the authors obtain forecasts of power line capacities to 2050, allowing only for increases in existing capacities without adding power lines between balancing areas. To be consistent with the Cambium database, the *Mid_ Case* scenario could be chosen.

For transmission efficiency, modeled by γ_{lt}, it is convenient to opt for the same methodology as the ReEDS model, i.e. to choose the convention of 1% loss for every 100 miles of electrical transmission line.

In order to implement this model with transmissions, we would still have to choose the LDES capacities for each state. Indeed, with a single state, we were varying the LDES capacity installed in the state, which modified the opportunity cost of the technology. But now, each state should have its own LDES capacity, and we need to set these to determine the opportunity costs. It would take too long to vary each LDES capacity in each state. To start with, we could, for example, build on the national case study in section 3 by choosing for each state the LDES capacity that gave the maximum opportunity cost to the technology. We could then compare the results and see whether the opportunity costs would differ from the model without transmissions.

4.4 Preliminary investigation of a time-sampling method

On the machine used, there wasn't enough memory to run the multi-state model. We then tried to investigate temporal sampling to reduce the computational cost.

In power systems optimization problems, there are several ways of making a temporal trade-off between accuracy and computational requirements. The two most commonly used are optimization on a subset of days representative of the year, and changing the temporal timestep (from hourly to every 2h, 4h, 8h... or variable duration timesteps) [16]. For our problem, it seems more interesting to opt for the second method, since the study concerns LDES, and we want to track its behavior, particularly its state of charge, over the whole year. In [17], the authors find a way to decrease temporal resolution in solar-dominated grids. They rely on the periodicity of solar production to identify, for each day, 2 timesteps instead of 24: one hour after sunrise and one hour before sunset. For capacity-expansion problems, they manage to stay below 10% error. We propose an adaptation of the current constraints to decrease temporal resolution.

In the model, we need to change some constraints and especially the ones that specify state of charge evolution. In this model, t indexes still represent timesteps, but are no longer necessarily one hour long. We name τ_t the number of hours of the time step t. Then, we have for the state of charge of a battery:

$$v_{ht} = v_{h,t-1} + \tau_t \left(\eta_h p_{ht}^{st,ch} - p_{ht}^{st,dis} \right)$$

and for the ramping constraints of generators:

$$\begin{cases} p_{g,t} - p_{g,t-1} \leq RU_g^{factor} P_g \tau_{t-1} \\ p_{g,t-1} - p_{g,t} \leq RD_g^{factor} P_g \tau_{t-1} \end{cases}$$

In the objective function as well, we want the temporal part to be weighted the same as without temporal sampling. To compensate for the smaller number of timesteps, we weight each temporal term by the corresponding number of time steps. For example, the objective function of the baseline becomes:

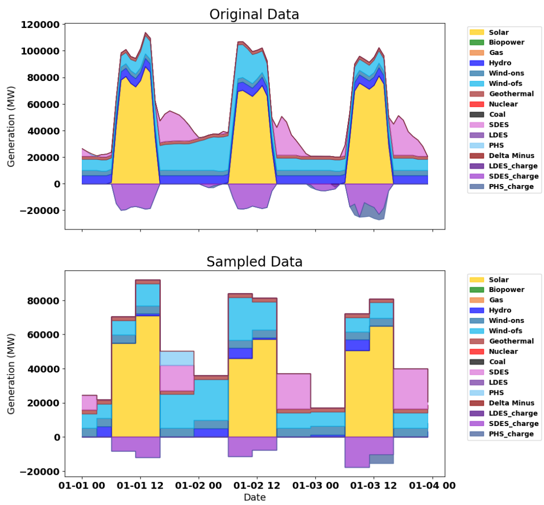
$$\begin{aligned} \underset{\substack{p_{gt}, p_{ht}^{st,ch}, p_{ht}^{st,dis}, \\ r_{gt}^{up}, \Delta_t^-, \Delta_t^+, \delta_t^{up,short}}}{\text{Minimize}} \quad & \sum_{t \in T} \Delta_t \left[\sum_{g \in G} \left[C_{gt}^{p} p_{gt} + C_{gt}^{up} r_{gt}^{up} \right] + C^I \left(\Delta_t^- + \Delta_t^+ \right) + C^{short} \delta_t^{up,short} \right] \\ & + \sum_{g \in G^{firm,fixed}} C_g^{fom,gen} \bar{p}_g + \sum_{g \in G^{renew,fixed}} C_g^{fom,gen} \bar{P}_g \\ & + \sum_{h \in H^{fixed}} C_h^{fom,st,power} \bar{P}_h^{st,power} \end{aligned}$$

All time values used as inputs must also be adapted: for demand, the power demand over a range τ_t will be equal to the average of the power demand over each hour of this range. For renewable capacity factors, the value for timestep t is also the average capacity factor for each hour in the range τ_t.

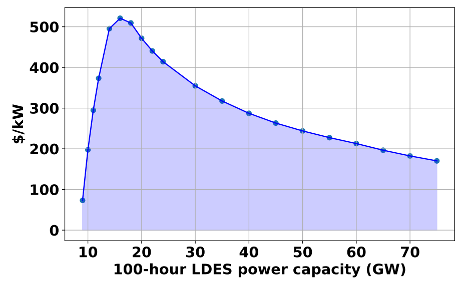
Even though we didn't have time to test this temporal sampling with the transmission model, it's still interesting to study its impact with an example. We tried out the state of California and compared temporal sampling with the hourly version, for the baseline model and also for the OVM model. We tested a number of sampling methods, and the method that was chosen was to take 4 timesteps per day: one at sunrise, one at sunset and two in the middle of these hours.

The result for the baseline is a difference of 0.16% in total costs and 0.85% in operation costs compared to the original California baseline model. However, for the opportunity value maximization model, we obtain significantly different results for boundary costs and corresponding capacity, as shown in figure 11b and 11c. We still have the same curve shape, but in the model with sampling the LDES capacity at which the problem becomes feasible is lower(4 GW instead of 9). The maximum boundary cost is obtained for 6 GW with temporal sampling instead of 17 GW for the hourly version.

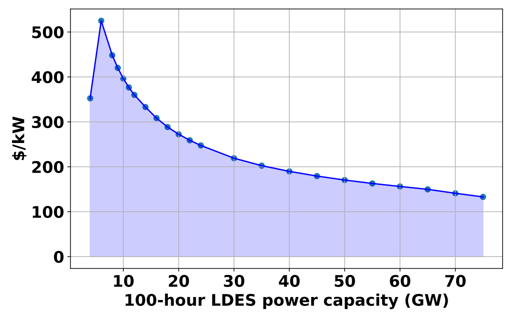
This could be due to the fact that demand is averaged over several hours, and therefore the maximum demand to be supplied is lower than that of the hourly-version (see 11a), which could lead to an underestimation of the capacities needed to meet demand. A method of reducing the number of variables where maximum demands are kept could probably help to correct this flaw.



(a) Hourly (top) and sampled (bottom) predicted generation and storage charge of baseline model for California from January 1 to 3, 2050



(b) Boudary costs of LDES in California with hourly temporal resolution



(c) Boudary costs of LDES in California with variable temporal resolution (4 timesteps a day)

Figure 11: Comparison of results for California state with and without temporal sampling

Conclusion

This exhaustive study of the 48 contiguous states of the United States highlights the great heterogeneity of the territories and the formidable challenge of making the transition to decarbonized energy using long-term energy storage. It will eventually be included in a report for decision-makers for the US Department of Energy (DoE). According to current results, with learning rate projections of LDES technology, only a small number of states could economically competitively replace gas and coal with long-term storage systems and investment in solar and wind generators only. To make this transition, subsidies will be needed, or the deployment of LDES will have to be considered in conjunction with other technical solutions.

There are still improvements to be made in this model, the most important being the introduction of uncertainty to obtain more robust results, and the consideration of interstate connections to obtain results on the opportunity cost of LDES technology in the USA. We have established the foundations of a model that takes transmissions into account. We also proposed a temporal sampling model that would reduce the computational cost of the model, although its effectiveness for the opportunity value maximization model is not yet satisfying. This will be taken up by Patricia Silva in a future study. Finally, we are largely basing our model on the forecasts obtained by NREL about capacity expansion, based on a number of assumptions about the evolution of the American power grid. It would be interesting to study other scenarios proposed by NREL to obtain a more complete view of the feasibility of LDES technologies.

References

[1] Paul Donohoo-Vallett, Nicole Ryan, and Ryan Wiser. *On the Path to 100% Clean Electricity*. U.S. Department of Energy's Office of Policy, May 2023. URL: https://www.energy.gov/policy/articles/path-100-clean-electricity.

[2] Patricia Silva et al. *Boundary Technology Costs for Economic Viability of Long-Duration Energy Storage Systems*. 2024. eprint: 2401.13163. URL: https://arxiv.org/abs/2401.13163.

[3] *The Long-term Strategy of the United States: Pathways to Net-zero Greenhouse Gas Emissions by 2050*. United States Department of State, 2021. URL: https://books.google.com/books?id=v8WWzwEACAAJ.

[4] Jeremy Twitchell, Kyle DeSomber, and Dhruv Bhatnagar. "Defining long duration energy storage". In: *Journal of Energy Storage* 60 (2023), p. 105787. ISSN: 2352-152X. DOI: https://doi.org/10.1016/j.est.2022.105787. URL: https://www.sciencedirect.com/science/article/pii/S2352152X22017753.

[5] Nestor Sepulveda et al. "The design space for long-duration energy storage in decarbonized power systems". In: *Nature Energy* 6 (May 2021), pp. 1–11. DOI: 10.1038/s41560-021-00796-8.

[6] Carlos Medina, C Ríos M Ana, and Guadalupe González. "Transmission grids to foster high penetration of large-scale variable renewable energy sources–A review of challenges, problems, and solutions". In: *International Journal of Renewable Energy Research (IJRER)* 12.1 (2022), pp. 146–169.

[7] Jonathan Ho et al. *Regional Energy Deployment System (ReEDS) Model Documentation: Version 2020*. Tech. rep. NREL/TP-6A20-78195. Golden, CO: National Renewable Energy Laboratory, 2021. URL: https://www.nrel.gov/docs/fy21osti/78195.pdf.

[8] Victor Diakov et al. *Improving Power System Modeling: A Tool to Link Capacity Expansion and Production Cost Models*. Tech. rep. National Renewable Energy Laboratory, 2021. URL: https://www.nrel.gov/docs/fy16osti/64905.pdf.

[9] P. Gagnon, B. Cowiestoll, and M. Schwarz. *Cambium 2022 Scenario Descriptions and Documentation*. Tech. rep. NREL/TP-6A40-84916. [Online]. Available: https://nrel.gov/publications.html. National Renewable Energy Laboratory (NREL), Jan. 2023.

[10] U.S. Energy Information Administration. *Annual Energy Outlook 2023*. Full Text: Download Resource. Mar. 2023. URL: https://www.eia.gov/outlooks/aeo/.

[11] LDES Council. *FLAGSHIP 2021 NET-ZERO POWER REPORT*. Exhibit 24 p. 29. 2021. URL: https://www.ldescouncil.com/assets/pdf/LDES-brochure-F3-HighRes.pdf.

[12] Scott Lundberg and Su-In Lee. *A Unified Approach to Interpreting Model Predictions*. 2017. arXiv: 1705.07874 [cs.AI]. URL: https://arxiv.org/abs/1705.07874.

[13] Jidong Kang et al. "A stochastic-robust optimization model for inter-regional power system planning". In: *European Journal of Operational Research* 310.3 (2023), pp. 1234–1248. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2023.03.024. URL: https://www.sciencedirect.com/science/article/pii/S037722172300231X.

[14] E Ela, M Milligan, and B Kirby. "Operating Reserves and Variable Generation". In: (Aug. 2011). DOI: 10.2172/1023095. URL: https://www.osti.gov/biblio/1023095.

[15] Pieter Gagnon et al. *2023 Standard Scenarios Report: A U.S. Electricity Sector Outlook*. Tech. rep. NREL/TP-6A40-87724. Golden, CO: National Renewable Energy Laboratory, 2023. URL: https://www.nrel.gov/docs/fy24osti/87724.pdf.

[16] Bethany A. Frew and Mark Z. Jacobson. "Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model". In: *Energy* 117 (2016), pp. 198–213. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2016.10.074. URL: https://www.sciencedirect.com/science/article/pii/S0360544216315110.

[17] Farzan ZareAfifi, Zabir Mahmud, and Sarah Kurtz. "Diurnal, physics-based strategy for computationally efficient capacity-expansion optimizations for solar-dominated grids". In: *Energy* 279 (2023), p. 128206. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2023.128206. URL: https://www.sciencedirect.com/science/article/pii/S0360544223016006.

A Additional constraints in baseline and OVM models

A.1 Storage devices

A.1.1 State of charge of storage systems

The operational constraints on storage systems are based on the state of charge, which must be the same at the first and last timestep of the simulation (10c) so as not to use more energy than is created, and then evolves according to the charging and discharging powers at each timestep(10b and 10a). The constraints 10d and 10e describe the extreme capacity constraints for firm and candidate storage systems respectively.

$$v_{ht} = v_{h,t^{ini}} + \eta_h p_{ht}^{st,ch} - p_{ht}^{st,dis}; \forall h \in H, t = 1 \tag{10a}$$
$$v_{ht} = v_{h,t-1} + \eta_h p_{ht}^{st,ch} - p_{ht}^{st,dis}; \forall h \in H, t \in T \mid t \geq 2 \tag{10b}$$
$$v_{h,t^{ini}} = v_{h,t^{end}}; \forall h \in H \tag{10c}$$
$$0 \leq v_{ht} \leq \bar{V}_h; \forall h \in H^{fixed}, t \in T \tag{10d}$$
$$0 \leq v_{ht} \leq x_h^{st,energy\dagger}; \forall h \in H^{cand}, t \in T \tag{10e}$$

A.1.2 Power and reserve constraints

Charging and discharging powers are also limited for each storage system, so as not to exceed the charging/discharging limits per timestep (11a and 11c), even when taking into account the storage systems' reserve (11b and 11d). Each storage system must be able to discharge the reserve it has built up at any time (11e).

$$0 \leq p_{ht}^{st,ch} \leq \bar{P}_h^{st,power}; \forall h \in H^{fixed}, t \in T \tag{11a}$$
$$0 \leq p_{ht}^{st,dis} + r_{ht}^{sl,up} \leq \bar{P}_h^{st,power}; \forall h \in H^{fixed}, t \in T \tag{11b}$$
$$0 \leq p_{ht}^{st,ch} \leq x_h^{st,power}; \forall h \in H^{cand}, t \in T \tag{11c}$$
$$0 \leq p_{ht}^{st,dis} + r_{ht}^{st,up} \leq x_h^{st,power}; \forall h \in H^{\text{cand}}, t \in T \tag{11d}$$
$$v_{ht} - r_{ht}^{st,up} \geq 0; \forall h \in H, t \in T \tag{11e}$$
$$p_{ht}^{sl,ch}, p_{ht}^{sl,dis}, r_{ht}^{sl,up} \geq 0; \forall h \in H, t \in T \tag{11f}$$

A.1.3 Capacity constraints

The maximum energy of each storage system is related to its rated power by its duration (12a. In the opportunity value maximization model, the maximum power rating is bounded by a maximum investment capacity (12b.

$$x_h^{st,energy\dagger} = x_h^{st,power\dagger}\bar{S}_h; \forall h \in H^{cand} \tag{12a}$$
$$0 \leq x_h^{st,power\dagger} \leq \bar{x}_h^{st,power}; \forall h \in H^{cand} \tag{12b}$$

A.2 Generators

A.2.1 Generator investment and retirement

Investment in new generators must be lower than a maximum capacity, which is expressed by the 13a constraint. In addition, in the opportunity value maximization model, some generators are retired, as indicated by the constraint 13b. The power production remaining to the retired generator g is p_g^{rem}. The value of retirement is determined by the minimum $\underline{x}_g^{ret,gen}$ and maximum $\bar{x}_g^{ret,gen}$ in the 13cconstraint.

$$0 \leq x_g^{inv,gen\dagger} \leq \bar{x}_g^{inv,gen}; \forall g \in G^{cand} \tag{13a}$$
$$\bar{p}_g^{rem} = \bar{P}_g - x_g^{ret,gen\dagger}; \forall g \in G^{firm,fixed} \tag{13b}$$
$$\underline{x}_g^{ret,gen}\bar{P}_g \leq x_g^{ret,gen\dagger} \leq \bar{x}_g^{ret,gen}\bar{P}_g, \forall g \in G^{firm,fixed} \tag{13c}$$

A.2.2 Production, reserve and capacity factor constraints

Each generator has a maximum production per timestep (14). In addition, every intermittent generator g (photovoltaic, wind, hydro) sees its production at each time t limited by the availability of the resource represented by $f_{gt}^{available}$, called capacity factor. For reserve providers generators, production must be such that even if the reserve has to be produced, the production and reserve power do not exceed the maximum production capacity (14b and 14e).

$$0 \leq p_{gt} \leq \bar{p}_g^{rem} - r_{gt}^{up}; \forall g \in G^{firm,fixed}, t \in T \tag{14a}$$
$$0 \leq p_{gt} \leq \bar{P}_g f_{gt}^{available}; \forall g \in G^{renew,fixed} \mid g \notin G^{res,providers}, t \in T \tag{14b}$$
$$0 \leq p_{gt} \leq \bar{P}_g f_{gt}^{available} - r_{gt}^{up}; \forall g \in G^{renew,fixed} \cap G^{res,providers}, t \in T \tag{14c}$$
$$0 \leq p_{gt} \leq x_g^{inv,gen\dagger} f_{gt}^{available}; \forall g \in G^{renew,cand} \mid g \notin G^{res,providers}, t \in T \tag{14d}$$
$$0 \leq p_{gt} \leq x_g^{inv,gen\dagger} f_{gt}^{available} - r_{gt}^{up}; \forall g \in G^{renew,cand} \cap G^{res,providers}, t \in T \tag{14e}$$

In a similar way to storage systems, each reserve provider generator g sees its reserve capacity limited by a maximum contribution corresponding to a fraction $R_g^{up,factor}$ of the maximum capacity it could produce at that timestep (15).

$$0 \leq r_{gt}^{up} \leq \bar{p}_g^{rem} \bar{R}_g^{up,factor}; \forall g \in G^{firm,fixed}, t \in T \tag{15a}$$
$$0 \leq r_{gt}^{up} \leq \bar{P}_g f_{gt}^{available} \bar{R}_g^{up,factor}; \forall g \in G^{renew,fixed} \cap G^{res,providers}, t \in T \tag{15b}$$
$$0 \leq r_{gt}^{up} \leq x_{gt}^{inv,gen\dagger} \bar{R}_g^{up,factor}; \forall g \in G^{firm,cand}, t \in T \tag{15c}$$
$$0 \leq r_{gt}^{up} \leq x_g^{inv,gen\dagger} f_{gt}^{available} \bar{R}_g^{up,factor}; \forall g \in G^{renew,cand} \cap G^{res,providers}, t \in T \tag{15d}$$
$$\tag{15e}$$

A.2.3 Ramp-up and ramp-down constraints

Each generator has properties which imply that production variations between timesteps are limited. These ramp-up and ramp-down constraints are represented as follows:

$$p_{gt} - p_{g,t-1} \leq RU_g^{factor} \bar{p}_g^{rem}; \forall g \in G^{firm,fixed}, t \in T \mid t \geq 2 \tag{16a}$$
$$p_{g,t-1} - p_{gt} \leq RD_g^{factor} \bar{p}_g^{rem}; \forall g \in G^{firm,fixed}, t \in T \mid t \geq 2 \tag{16b}$$
$$p_{gt} - p_{g,t-1} \leq RU_g^{factor} x_g^{inv,gen\dagger}; \forall g \in G^{firm,cand}, t \in T \mid t \geq 2 \tag{16c}$$
$$p_{g,t-1} - p_{gt} \leq RD_g^{factor} x_g^{inv,gen\dagger}; \forall g \in G^{firm,cand}, t \in T \mid t \geq 2 \tag{16d}$$

A.3 Setup in baseline and opportunity value maximization models

In the baseline model, neither investment nor technology retirement is authorized in constraints 12b, 13a and 13c:

$$\begin{aligned}
&\underline{x}_g^{ret,gen} = \bar{x}_g^{ret,gen} = 0 \quad \forall g \in G^{firm,fixed} \\
&\bar{x}_g^{inv,gen} = 0 \quad \forall g \in G^{cand} \\
&\bar{x}_h^{st,power} = 0 \quad \forall h \in H^{cand}
\end{aligned}$$

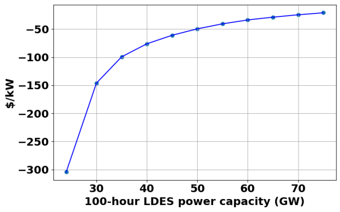
In the opportunity value maximization model, gas and coal generators are retired. Renewable candidate generators are considered, as well as storage systems. We then have:

$$\begin{aligned}
&\underline{x}_g^{ret,gen} = \bar{x}_g^{ret,gen} = 1 \quad \forall g \in G^{gas,fixed} \cup G^{coal,fixed} \\
&\underline{x}_g^{ret,gen} = \bar{x}_g^{ret,gen} = 0 \quad \forall g \in G^{firm,fixed} \setminus \left(G^{gas,fixed} \cup G^{coal,fixed}\right) \\
&\bar{x}_g^{inv,gen} \geq 0 \quad \forall g \in G^{cand} \\
&\bar{x}_h^{st,power} \geq 0 \quad \forall h \in H^{cand}
\end{aligned}$$

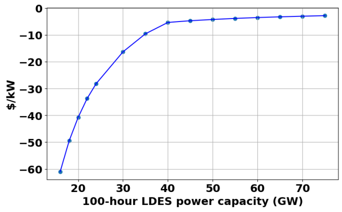
Also, for LDES systems, we set their installed capacity in the OVM model, which means adding the following constraint:

$$x_h^{st,power} = x_h^{st,power\ddagger}, \forall h \in H^{cand,long} \tag{19}$$

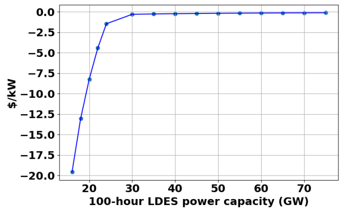
B Negative boundary costs



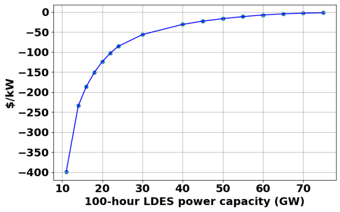
(a) Alabama negative cost



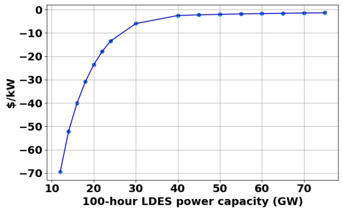
(b) Connecticut negative cost



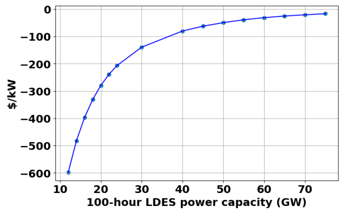
(c) Delaware negative cost



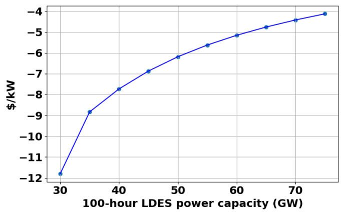
(d) Florida negative cost



(e) New Jersey negative cost



(f) Ohio negative cost



(g) Louisiana negative cost

Figure 12: Negative boundary costs

C List of acronyms used in the report

LDES Long-Duration Energy Storage

SDES Long-Duration Energy Storage

OVM Opportunity Value Maximization

BA Balancing Area