Received: 28 February 2020 | Revised: 21 April 2020 | Accepted: 27 April 2020
DOI: 10.1002/er.5551

RESEARCH ARTICLE

INTERNATIONAL JOURNAL OF ENERGY RESEARCH WILEY

# Reinforcement learning for electricity dispatch in grids with high intermittent generation and energy storage systems: A case study for the Brazilian grid

**Vinícius de Carvalho Neiva Pinheiro**[1] | **Alberto L. Francato**[1] | **Warren B. Powell**[2]

[1]Department of Water Resources, School of Civil Engineering, Architecture and Urban Design, University of Campinas (UNICAMP), Campinas, Brazil

[2]Department of Operations Research and Financial Engineering, Princeton University, Princeton, New Jersey

**Correspondence**
Vinícius de Carvalho Neiva Pinheiro, Department of Water Resources, School of Civil Engineering, Architecture and Urban Design, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.
Email: pinheiro@fec.unicamp.br

**Funding information**
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

**Summary**

Intermittent energy sources such as wind and solar have recently been growing a lot faster than dispatchable energy sources in Brazil, which made investments in energy storage systems become an attractive possibility in the country. Current operational policies for energy dispatch do not consider storage systems and need adjustments to fit this technology. With this motivation, we use reinforcement learning techniques to develop policies for managing storage systems in a grid that can handle time-varying inputs and loads, with rolling forecasts. We use a deterministic lookahead (DLA) policy which has been parametrically modified to perform well in the presence of uncertain forecasts. For realistic simulations, the base model considers important characteristics in a grid that influence the interaction between scheduling and real-time operation such as power and ramping capacities, notification times, and stochastic forecasts. The parametric modification with tunable parameters allows an optimal balance between two conflicting services provided by the storage system: time-shifting and spinning reserves. Optimal reserves ranged from 35% to 100%, depending on the tested dataset, which shows the importance of tuning. Differently from stochastic lookahead policies, which are computationally expensive, parameterized DLA policies can be applied to real-time operation after being optimized in a stochastic base model.

**KEYWORDS**

energy storage, intermittent energy sources, operational policies, policy design, stochastic forecast

## 1 | INTRODUCTION

Clean and renewable energy sources, which include wind, solar photovoltaic, geothermal, hydro, biofuel, biomass, and tidal, are rapidly growing their participation in worldwide power capacity.[1] The Renewables Global Status Report 2019 points out that wind power capacity in the world expanded from 540 GW in 2017 to 591 GW in 2018: an increase of 9.4%. In the same period, solar photovoltaic power capacity increased 24.7%, from 405 to 505 GW.[2]

In accordance with global trends, wind and solar sources have been growing rapidly in Brazil and are quickly changing the energy mix of the country.[3] Most of Brazilian electricity generation capacity comes from hydro. However, while it represented more than 90% of the national capacity 20 years ago, it is now less than 70%. According to the Brazilian National Grid Operator, in August 2008, hydroelectric generation represented 83.2% of the total capacity and, in August 2018, this percentage was 67.8%. In addition, while there was virtually no on-grid solar or wind capacity in 2008, by 2018, these sources had grown to 0.8% and 8.3%, respectively (Table 1).[4]

Wind and solar feature low environmental impacts, and can be economically attractive in certain conditions. These features are going to drive their growth for some time. However, wind and solar introduce significant variability, as well as uncertainty, into the generation of energy, complicating the problem of meeting power needs with high reliability.[5-7]

Unlike hydroelectric plants, which allow regulation of power output by controlling water flow through turbines, generation from wind farms depends on wind velocities, which are not only highly variable but also are hard to forecast.[8] Solar farms generation is also controlled by an exogenous uncontrolled process (solar radiation). For very sunny or very cloudy days, its prediction can be considerably more precise than wind generation prediction. However, days with spotted clouds are particularly hard. They can introduce very sudden jumps in power, and they are hard to predict even an hour into the future.[9]

To make generation meet demand, dispatchable power plants must be capable of generating the difference between demand and intermittent generation. Hydro and thermal both represent dispatchable energy sources for the grid, but not all of them have the ability to mitigate the intermittency of renewables. The issue here is notification times and ramping.[10] Power sources which are able to mitigate intermittency, also called load-following plants, must be able to vary their output with sufficient speed and within short enough notice.[11,12]

Concerning dispatchability, hydropower plants can be divided into two categories: those with and those without reservoirs.[13] Those with reservoirs have the ability to vary the level of their lakes, storing water to use in the future.

**TABLE 1** Brazilian energy mix in 2009 and 2019

| | May, 2009 | May, 2019 |
|---|---|---|
| | Capacity (GW) | Capacity (GW) |
| Hydro | 83.2 (82%) | 110.4 (67%) |
| Thermo | 16.5 (16%) | 34.2 (21%) |
| Nuclear | 2.0 (2%) | 2.0 (1%) |
| Wind | 0.3 (0%) | 15.0 (9%) |
| Solar | 0.0 (0%) | 2.1 (1%) |
| Total | 102.0 | 163.7 |

*Source*: Adapted from ONS.[4]

Hydro plants without reservoirs (also known as run-of-the-river hydro plants) are constrained by the instantaneous incoming river flow, which limits the ability to adjust their power output. In Brazil, hydro plants are still the main source of electricity (68% of total), but, for environmental reasons, run-of-the-river hydro plants have been favored for new investments, which require less flooded areas but are not dispatchable.[8] Therefore, storage capacity relative to demand has been decreasing in the country (Figure 1).[14]

Thermal plants can be divided into two categories with respect to dispatchability: gas turbine and steam. While the first type usually has fast ramping capacities, the latter can take hours to start up and its output variation is usually slow.[15] Brazil has not experienced outages due to sudden drops in wind or solar generation because it has enough capacity of fast ramping thermal plants to mitigate its current wind and solar capacity. However, the generation of these thermal plants is relatively expensive.[16]

The Brazilian Interconnected Grid covers 98.3% of the demand for electricity in the country and is divided in four large subsystems connected by transmission lines: South, Southeast/CentralWest, Northeast, and North.[17] The greatest part of the Brazilian intermittent installed capacity is located in the Northeast. 84% of the installed wind generation capacity is in this subsystem, along with 67% of the solar. The Northeast has only 10% of the Brazilian hydro capacity.[4] The energy mix of the Northeast subsystem is represented in Table 2.

Wind farms are responsible for more than 40% of the Northeast generation. In this subsystem, wind intermittency is compensated by "imported" energy from other subsystems. Energy import and export between subsystems, called energy exchange, is possible through very long distance transmissions, which allows surplus generation from one subsystem to complement generation from another, but incurring transmission losses. In Brazil, it is estimated that transmission losses consume around 5% of total electricity production and, as that number is the country's average percentage, it is reasonable to assume that energy exchange losses are higher than that, as these transmissions involve much longer distances than the transmissions within the subsystems. Moreover, the grid operator is concerned about overloading the existing transmission lines,[17] as new ones are not being built in a sufficient pace to keep up with the increasing energy exchange needs.

The Northeastern subsystem is primarily an importer of energy. Even though it exports surplus wind generation for around 70 days a year, in 2017, total energy imports were 10 times greater than total exports.[4] Figure 2 shows generation data from hydro, thermal, and

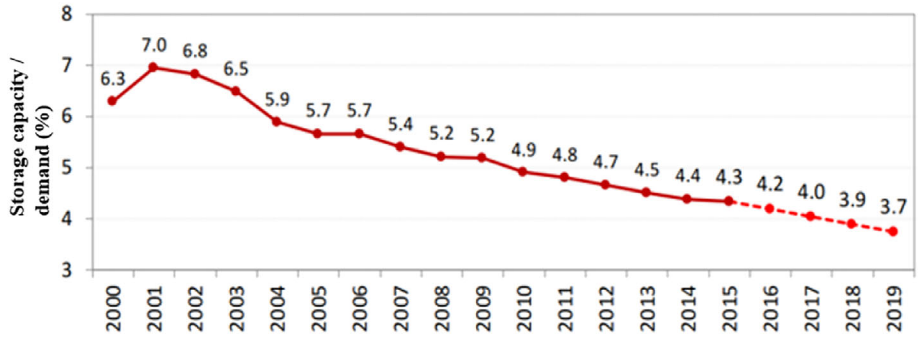


**FIGURE 1** Evolution of storage capacity relative to demand in Brazil[14] [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 2** Northeastern energy mix

| | May, 2019 | |
|---|---|---|
| | Capacity (GW) | % of subsystem |
| Hydro | 11.0 | 34.2 |
| Thermal | 7.2 | 22.4 |
| Nuclear | 0.0 | 0.0 |
| Wind | 12.6 | 39.1 |
| Solar | 1.4 | 4.3 |
| Total | 32.2 | 100.0 |

*Source*: Adapted from ONS.[4]

wind plants in the Northeast subsystem for a 40 day period. It can be observed in the graph how wind intermittence is compensated by energy exchange at times when wind generation is relatively low.

Just as Brazil is experiencing growth in renewables, it has to manage both the variability and uncertainty of wind and solar with the continuing need to meet their peak demands. In addition, storage capacity (in the form of hydroelectric reservoirs) is not growing with demand,[14] which means that it is shrinking as a percentage of total generation capacity, introducing limits on the ability of the system to handle variability. This situation indicates that energy storage systems can be useful solutions to increase the economic efficiency of the Brazilian electricity sector operation, as these systems are reliable and widely tested in other countries for mitigation of wind and solar intermittency, while also improving dispatchability and storage capacity in an electricity system.

Because of its predominantly hydro generation capacity, the Brazilian grid has not experienced a need for energy storage until the recent growth of intermittent sources and, therefore, the Brazilian grid operator does not have experience with energy storage systems operation nor an official dispatch model that considers this kind of systems. However, the need for storage systems has been cited for the first time in the 2017 Brazilian Decennial Expansion Plan,[18] a document with guidelines for the development of the Brazilian electricity sector infrastructure.

Another necessity that intermittent renewables growth has brought to Brazil is the ability to deal with short-term uncertainty. As demand is sufficiently predictable, before wind capacity became significant, the Brazilian grid operator had to be concerned with the stochasticity related only to rain forecasts, which determine hydro generation availability. However, hydro availability is still predictable for the next day and rain forecasts are more useful for a longer term plan. On the other hand, wind generation brings uncertainty for the next day or even for the next few hours of planning the operation of a grid. Short-term wind forecasting is hard because we do not know the state of the atmosphere. Current short time models (up to 3-4 hours) tend to use "persistence forecasting," which means we assume that the wind will not change. This brings a big challenge for short-term planning, as a small drop in wind speed is enough to produce a rapid drop in power moves (roughly with the cube) to create an outage.

There are many uses of storage systems on the grid. Castillo and Gayme[19] list as grid-scale storage systems applications: spinning reserves, time-shifting (or energy-shifting), power quality, transient stability, regulation, voltage control, load following, firm capacity, congestion relief, and upgrade deferral. In this work, we focus on spinning reserve and time-shifting, the most common applications in grid-scale.[20]

Energy storage systems can bring great benefits to electricity grids, but only if operated appropriately.[21-23] Finding the best policy to operate those systems can be complicated as it depends on characteristics that can be very specific for some grids. For instance, demand patterns, energy mix, and spot market prices certainly have a big influence not only in the choice of storage system types, but also in their operational policies. Moreover, the uncertainty involving solar and wind farms generation brings an extra challenge to the development of efficient policies to operate electricity grids.

Studies involving development of policies to make sequential decisions are included in a vast research field

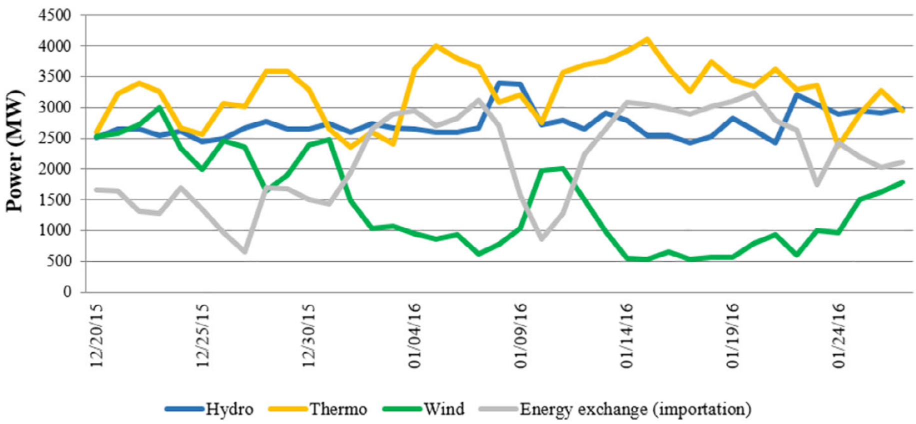


**FIGURE 2** Energy exchange and generation from different sources in Northeast (adapted from Reference 8) [Colour figure can be viewed at wileyonlinelibrary.com]

called Reinforcement Learning. The term Reinforcement Learning, however, is still not well defined in the research community. It started in the 1990's with Q-Learning, but evolved to other proposed methods, as Q-Learning often did not work for some problems.

There are many articles that use reinforcement learning to propose operational policies for storage systems, but typically ignore the presence of rolling forecasts.[24-27] Some authors consider forecasts, but use them in deterministic, rolling-horizon models which ignore the uncertainty in the forecasts.[28,29] Arnold and Andersson (2011) proposes a model predictive control technique to schedule storage systems operation. To deal with forecasts errors, they set reserves, but the reserves are predefined and the storage systems are only optimized for time-shifting.[30]

Joint optimization of time-shifting and reserves for storage systems has been explored in References 31-33 and, 34 but they do not consider a grid with multiple sources of energy with different operational constraints (ramping capacities, power, and notification time) and whose generation has to be scheduled. A simulator for a large-grid that addresses these issues, called SMART-ISO, was presented in References 35,36. This two-part study models the effects of large penetrations of offshore wind power into a large electric system using realistic wind power forecast errors and a complete model of unit commitment. Lu et al[37] use two-stage stochastic programming with scenario trees to handle the uncertainty of the future. This produces large models that are computationally expensive.

The computational burden of stochastic programs increases rapidly with the number of scenarios and variables, and horizon length.[38,39] In this article, we use a method that handles uncertainty in the future, without the computational overhead of stochastic programming. Since the parameters are tuned in a realistic simulator with real data from the Brazilian grid, the parameterized lookahead model is optimized under very realistic conditions. Moreover, in this work, we address a specific issue of the Brazilian grid, which traditionally used hydro for spinning reserves, but has an increasing need for more reserve sources as wind and solar capacities grows.

Introducing storage systems to an existing grid may require a careful reevaluation of its operational policies to fully benefit from the features of these systems.[40] With this motivation, we present in this work a dispatch model, in the form of a simulator, to evaluate and compare different multistage policies to operate grids with storage systems and high participation of intermittent sources. The policies simulate planning tools similar to the ones already used in Brazil for generation planning and real-time operation, but with modifications to increase the ability to deal with uncertainty and storage systems. A lookahead (also known as Model Predictive Control in the controls community) is run hourly and it considers a wind generation forecast that is updated in every run, while real-time operation is decided using a parametric decision rule that has been optimized using the simulator. Different parametric modifications are used in the lookaheads for each policy, which changes the role and behavior of the storage systems and hydros.

In one of the policies that we test, we assume that storage systems can only provide time-shifting, while reserves are provided only by hydro. In a second policy, reserves are only provided by storage systems, which has its time-shifting capability reduced as reserve needs grows. In a third policy, both storage systems and hydro can provide reserves and the amounts of reserves for each source are mutually optimized. Results show that the third policy, in which hydro and storage reserves are jointly optimized while still allowing storage systems to provide time-shifting, have significantly better performances than the policies that only allow one source to provide reserves.

The main contribution of this article is the development of a novel technique that can handle energy storage systems and stochastic rolling forecasts, which cannot be found elsewhere in the literature. Instead of using stochastic lookahead techniques, which are computationally expensive and inapplicable to real-time operation, we use a parameterized deterministic lookahead (DLA) that is optimized in a stochastic base model. Both real-time and planning policies are tuned with a very realistically modeled grid. Another important contribution is a policy that allows time-shifting and spinning reserve to be jointly optimized for storage systems.

The article is organized as follows. In Section 2, we present the base model, which is a detailed simulator of the Northeast of Brazil used to evaluate different storage management policies. In Section 3, we present a general framework for designing policies, which includes four fundamental classes of policies. In Sections 4 to 6, we design different policies, focusing primarily on the principle of parameterized cost function approximations (CFAs) that combine a deterministic forecast with tunable parameters to improve the performance under uncertainty. Section 7 describes how the uncertainty involving the problem is modeled. In Section 8, we describe the optimization processes of the tunable parameters. In Section 9, we present and compare the results. Section 10 concludes the article.

## 2 | THE BASE MODEL

We first present the base model in the form of a simulator to test policies for managing energy generation. The model requires that we represent the planning process over time, which includes making advance commitments as necessary. We also need to model the different forms of uncertainty that enter our model as exogenous information. Included in our aggregate model will be hourly forecasts of wind, solar generation, and demand, as well as generation costs. Decisions for each time-step are hydro and thermal generation, imported or exported energy from or to other subsystems, and energy flow from or to the storage system.

Some approximations are made to reduce the complexity of the problem. For instance, the model considers only one subsystem of the country, the Northeast, which is the one with the highest wind and solar capacity. Moreover, the generation units are modeled as aggregate and six sources of energy are considered: wind, solar, hydro, thermal, storage system, and energy imports. We also do not model the transmission grid, and hence transmission losses and constraints are not considered.

Input data for the model include operational characteristics of hydro, thermal, and energy storage system; forecasts for wind and solar generation; and a demand forecast. Thermal generation and energy exchange marginal costs are also input data.

The thermal plants were divided into two categories: slow and fast ramping. Usually, these plants are simply divided into steam and gas turbine to represent the difference in their ramping capacities. However, we found it would be more accurate to use the actual ramping capacities, provided by the Brazilian Chamber of Electric Energy Commercialization to categorize them. This resulted in a group of slow ramping thermals which total of 2.5 GW of installed capacity and a group of fast ramping thermals with 4.8 GW. From now on, we will call these groups slow and fast thermals.

Five elements compose our model and are described in this section: state variables, decision variables, exogenous information variables, the transition function (that captures the evolution of the state variables over time), and the objective function, which is how we evaluate the performance of different policies.

### 2.1 | State variables

The dynamic state variable $S_t$ contains all the information needed to model the system from time $t$ onward,[41] while the initial state $S_0$ includes the initial values of all dynamically changing variables, or fixed parameters that do not change over time.

For $t = 0$, the observed actual wind generation $P_0^E$ is included in the initial state variable. We also include in $S_0$ the forecasts of demand $f_t^D$ and solar generation $f_t^{PV}$ for the whole simulation period, which are treated as perfect and static in the model and, thus are latent variables. There are forecasts in 15 minute increments over a 24 hour horizon, which means $H = 96$ time periods.

The initial state variable also contains: the vector $P^{\text{cap}}$ containing power capacities and minimum generation; the vector $\rho^{\text{cap}}$ containing ramping capacities; prices for energy exchange $c_t^{\text{EE}}$, deficit $c_t^{\text{DF}}$, slow thermal $c^{\text{ST}}(x^{\text{ST}})$, and fast thermal $c^{\text{FT}}(x^{\text{FT}})$; self-discharge $\eta^{\text{SD}} \in [0, 1]$, round-trip efficiency $\eta^{\text{RT}} \in [0, 1]$, maximum storage $R_{\max}^S$, and initial storage $R_0^S$ for the energy storage system; and initial available energy generation for hydro $R_0^{\text{HD}}$, which will be further discussed.

$$S_0 = \left( \left\{ f_t^{\text{PV}}, f_t^D, c_t^{\text{EE}}, c_t^{\text{DF}} \right\}_{t=0}^{T}, P_0^E, P^{\text{cap}}, \rho^{\text{cap}}, \eta^{\text{SD}}, \eta^{\text{RT}}, c^{\text{ST}}\left(x^{\text{ST}}\right), c^{\text{FT}}\left(x^{\text{FT}}\right), R_0^{\text{HD}}, R_{\max}^S, R_0^S \right), \tag{1}$$

$$P^{\text{cap}} = \left(P_{\text{max}}^{\text{HD}}, P_{\text{max}}^{\text{ST}}, P_{\text{max}}^{\text{FT}}, P_{\text{max}}^{\text{SD}}, P_{\text{max}}^{\text{SC}}, P_{\text{max}}^{\text{EE}}, P_{\text{min}}^{\text{HD}}, P_{\text{min}}^{\text{ST}}, P_{\text{min}}^{\text{FT}}, P_{\text{min}}^{\text{SD}}, P_{\text{min}}^{\text{SC}}, P_{\text{min}}^{\text{EE}}\right), \tag{2}$$

$$\rho^{\text{cap}} = \left(\rho_{\text{max}}^{\text{HD}}, \rho_{\text{max}}^{\text{ST}}, \rho_{\text{max}}^{\text{FT}}, \rho_{\text{max}}^{\text{SD}}, \rho_{\text{max}}^{\text{SC}}, \rho_{\text{max}}^{\text{EE}}, \rho_{\text{min}}^{\text{HD}}, \rho_{\text{min}}^{\text{ST}}, \rho_{\text{min}}^{\text{FT}}, \rho_{\text{min}}^{\text{SD}}, \rho_{\text{min}}^{\text{SC}}, \rho_{\text{min}}^{\text{EE}}\right). \tag{3}$$

Demand and solar forecasts are considered perfect here, and wind forecasts are updated in every time-step $t$. For $t > 0$, the wind forecasts are in the dynamic state variable, together with stored energy. We also include in the dynamic state variable decisions that were made before time $t$, to be implemented at times $t' \geq t$, which are commitments that have to be made in advance because of notification time constraints. The way the state variables evolve is described in Section 2.4 (Transition function) and in Section 7 (Modeling uncertainty).

$$S_t = \left(P_t^E, \left(f_{tt'}^E\right)_{t' \geq t}, R_t^S, \{x_{t-\tau,t'}\}_{t'=t+\tau}^{t+\tau+3}\right), \tag{4}$$

where $\tau$ is the notification time, which will be defined in Section 4, and $P_t^E$ is actual wind generation.

## 2.2 | Decision variables

For each time-step, the decision variables for the model are represented by $x_t$:

$$x_t = \left(x_t^{HD}, x_t^{ST}, x_t^{FT}, x_t^{SD}, x_t^{SC}, x_t^{EE}, x_t^{DF}\right), \tag{5}$$

which are, respectively, hydro generation, slow thermal generation, fast thermal generation, energy discharged from the storage system, energy charged to the storage system, energy imported from or exported to other subsystems (energy exchange), and energy deficit. We consider that wind and solar generation cannot be curtailed and, therefore, do not belong to the decision variable.

Decisions must respect the constraints in Equations (6) to (19). Total energy generation plus deficit in the grid must be equal or greater than total consumption, for each time-step:

$$x_t^{HD} + x_t^{ST} + x_t^{FT} + x_t^{SD} + x_t^{EE} + x_t^{DF} - x_t^{SC} \geq D_t - P_t^E - P_t^{PV}, \tag{6}$$

where $D_t$ and $P_t^{\text{PV}}$ are, respectively, observed demand and solar generation in time $t$. Because demand and solar generation are considered deterministic, then $P_t^D = f_t^D$ and $P_t^{\text{PV}} = f_t^{\text{PV}}$.

Energy generation from each source, in each time-step, must not exceed its capacity:

$$P_{min}^{HD} \leq x_t^{HD} \leq P_{max}^{HD}, \tag{7}$$

$$P_{min}^{ST} \leq x_t^{ST} \leq P_{max}^{ST}, \tag{8}$$

$$P_{min}^{FT} \leq x_t^{FT} \leq P_{max}^{FT}, \tag{9}$$

$$P_{min}^{SD} \leq x_t^{SD} \leq P_{max}^{SD}, \tag{10}$$

$$P_{min}^{SC} \leq x_t^{SC} \leq P_{max}^{SC}, \tag{11}$$

$$P_{min}^{EE} \leq x_t^{EE} \leq P_{max}^{EE}, \tag{12}$$

where $P_{max}^{HD}$, $P_{max}^{ST}$, $P_{max}^{FT}$, $P_{max}^{SD}$, and $P_{max}^{EE}$ are, respectively, the available generation capacity of hydro, slow thermal, fast thermal, storage system, and energy imports, and $P_{\text{max}}^{\text{SC}}$ is the charging capacity of the storage system. A minimum generation for each source is also set to approximate a constraint of minimum time on, present in some generation units, and spinning reserves.

Ramping capacities are also considered in the model. Depending on the type of generation unit, ramping capacity is not a constant value. It depends on the instantaneous output of the unit and whether it is ramping up or down. In this version of the model, ramping capacities are considered constant regardless of the output.

$$\rho_{min}^{HD} \leq x_{t+1}^{HD} - x_t^{HD} \leq \rho_{max}^{HD}, \tag{13}$$

$$\rho_{min}^{T} \leq x_{t+1}^{ST} - x_t^{T} \leq \rho_{max}^{ST}, \tag{14}$$

$$\rho_{min}^{T} \leq x_{t+1}^{FT} - x_t^{T} \leq \rho_{\text{max}}^{FT}, \tag{15}$$

$$\rho_{min}^{SD} \leq x_{t+1}^{SD} - x_t^{SD} \leq \rho_{max}^{SD}, \tag{16}$$

$$\rho_{min}^{SC} \leq x_{t+1}^{SC} - x_t^{SC} \leq \rho_{max}^{SC}, \tag{17}$$

$$\rho_{min}^{EE} \le x_{t+1}^{EE} - x_t^{EE} \le \rho_{max}^{EE}. \tag{18}$$

$\rho_{min}^{HD}$, $\rho_{max}^{HD}$, $\rho_{min}^{ST}$, $\rho_{max}^{ST}$, $\rho_{min}^{FT}$, $\rho_{max}^{FT}$, $\rho_{min}^{SD}$, $\rho_{max}^{SD}$, $\rho_{min}^{SC}$, $\rho_{max}^{SC}$, $\rho_{min}^{EE}$, and $\rho_{max}^{EE}$ are the minimum and maximum ramps for, respectively, hydro generation, slow thermal generation, fast thermal generation, storage system discharge, storage system charge, and energy exchange.

The storage system capacity constraint is represented by:

$$0 \le R_t^S \le R_{max}^S. \tag{19}$$

In the optimization process, a policy $X^{\pi}(S_t)$ produces a decision $x_t$ that satisfies all the constraints. Three policies are tested for the base model. In all policies, some decisions are made in advance of the time when they will be implemented. The results of each proposed policy are compared with a known-in-advance policy, in which wind forecasts are perfect. More details on the policies will be presented in Sections 4 to 6.

## 2.3 | Exogenous information

Changes in actual wind generation belongs to the exogenous information, which also includes the changes in forecasts for wind generation.

$$W_{t+1} = \left(\hat{P}_{t+1}^E, \left\{\hat{f}_{t+1,t'}^E\right\}_{t'=t+1}^{t+1+H} \forall t \in \mathcal{T}\right). \tag{20}$$

New information $W_{t+1}$ arrives every hour (every four time-steps of 15 minutes), or at times $t+1 \in \mathcal{T}$ (defined in Equation (21)). The policies are detailed in Section 3, but we anticipate that all policies are each composed of a combination of two policies, which are run with different frequencies: hourly and real time. That means that the hourly policy is run in every time $t \in \mathcal{T}$ and the real-time policy is run in every time-step $t$.

$$\mathcal{T} = \{0, 4, 8 \dots T\}, \tag{21}$$

The differences between wind forecasts and actual are uncertain and were modeled as presented in Section 7.

## 2.4 | Transition functions

The transition function determines how the state variables are updated when the system moves from $S_t$ to $S_{t+1}$. It can be represented as:

$$S_{t+1} = S^M(S_t, x_t, W_{t+1}). \tag{22}$$

Wind generation is updated according to:

$$P_{t+1}^E = P_t^E + \hat{P}_{t+1}^E. \tag{23}$$

Wind forecasts are updated in every time-step for $t \in \mathcal{T}$:

$$f_{t+4,t'}^E = f_{t,t'}^E + \hat{f}_{t+4,t'}^E, \forall t \in \mathcal{T}, \tag{24}$$

where $\hat{f}_{t+4,t'}^E$ is the change in the wind forecasts.

Energy stored in time $t+1$ depends on charge, discharge, round-trip efficiency, and energy lost due to self-discharge in time $t$. The energy storage equation is given by:

$$R_{t+1}^S = R_t^S - \left(R_t^S \eta^{SD}\right) + x_t^{SC} \eta^{RT} - x_t^{SD}. \tag{25}$$

Scheduled generation, which was previously decided in time $t - \tau$ has to be updated:

$$\{x_{t+1-\tau,t'}\}_{t'=t+\tau}^{t+\tau+3} = \{x_{t-\tau,t'}\}_{t'=t+1+\tau}^{t+\tau+4}. \tag{26}$$

## 2.5 | Objective function

The contribution function is composed by costs of thermal generation, energy exchange from or to other subsystems and deficit:

$$C(S_t, x_t) = c^{ST}\left(x^{ST}\right) x_t^{ST} + c^{FT}\left(x^{FT}\right) x_t^{FT} + c_t^{EE} x_t^{EE} + c_t^{DF} x_t^{DF}. \tag{27}$$

Note that energy exchange costs can be negative, as exports are allowed.

Both thermal and energy exchange prices are deterministic. Each one of the eight slow thermal plants and 33 fast thermal plants in the modeled system has its own generation cost. However, because thermal plants were modeled as aggregate, two quadratic functions were adjusted to determine marginal costs for thermal:

$$c^{ST}\left(x^{ST}\right) = \theta_0^{ST} + \theta_1^{ST} x^{ST} + \theta_2^{ST}\left(x^{ST}\right)^2, \tag{28}$$

$$c^{FT}\left(x^{FT}\right) = \theta_0^{FT} + \theta_1^{FT} x^{FT} + \theta_2^{FT}\left(x^{FT}\right)^2. \tag{29}$$

For energy exchange prices, we consider the average of spot market prices for all subsystems in Brazil other than the Northeast, weighted by their installed generation capacity. These prices are updated weekly in Brazil.

The objective function of the model looks for the best policy $\pi$ in a set of policies $\Pi$ to plan the operation of the modeled system with the lowest expected costs:

$$F = \min_{\pi \in \Pi} \mathbb{E}\left\{ \sum_{t=0}^{T} C(S_t, X^{\pi}(S_t)) | S_0 \right\}, \tag{30}$$

where

$$S_{t+1} = S^M(S_t, X^{\pi}(S_t), W_{t+1}). \tag{31}$$

In Sections 4 to 6, we present the policies $X^{\pi}(S_t)$ that we test to plan the operation of the system.

## 3 | DESIGNING POLICIES

Making decisions under uncertainty requires developing policies that map a state to a feasible decision. In other words, a policy defines the way a reinforcement learning agent behaves.[42,43] According to Powell[44], these policies can be divided into four fundamental classes:

- Policy function approximation (PFA)—These are analytical functions that directly map a state to an action, without any embedded optimization problem. These policies can be discrete lookup tables, parametric functions, or statistical models. An example of a parametric PFA, where a set of parameters $\theta$ would have to be tuned, could be:

$$X_t^{\pi}(S_t|\theta) = \theta_0 + \theta_1 S_t + \theta_2 S_t^2. \tag{32}$$

- Cost function approximation—These are parametric modifications of a cost function or constraints. The cost function can be altered by, for example, adding an error-correction term to it. The general form to represent a CFA is represented by the equation:

$$X_t^{\pi}(S_t|\theta) = \underset{x \in \mathcal{X}_t^{\pi}(\theta)}{\text{argmin}} \left( \bar{C}^{\pi}(S_t, x|\theta) \right). \tag{33}$$

- Value function approximation—Here, a function approximates the value or cost of being in a estate as a result of an action made in the previous time-step. This class of policy is also generally known as approximate dynamic programming and can be written as:

$$X_t^{\pi}(S_t|\theta) = \underset{x \in \mathcal{X}_t}{\text{argmin}} (C(S_t, x) + \mathbb{E}\{\bar{V}_{t+1}(S_{t+1}|\theta)|S_t\}), \tag{34}$$

where $S_{t+1}$ is updated by a transition function.

- Direct lookahead (DLA)—It solves an optimization lookahead problem over a horizon $H$. A DLA policy solves this problem deterministically, which, for instance, can be modeled as a linear program. The decision variables in the lookahead policy are represented with a tilde ($\tilde{x}_{tt'}$) not to be confused with the decision variables from the base model. A DLA can be represented as:

$$X_t^{\pi}(S_t|\theta) = \underset{\tilde{x}_{tt}, \ldots, \tilde{x}_{t,t+H}}{\text{argmin}} \sum_{t'=t}^{t+H} C\left(\tilde{S}_{tt'} \tilde{x}_{tt'}\right). \tag{35}$$

It is also possible to combine more than one class of policy to make hybrid ones. One example is using parametric CFAs in lookahead models, where modifications in a DLA (made by the addition of parameters in the objective function or constraints) are tuned in a stochastic simulator.

Three of the four classes of policies are employed in this work. The example of hybrid policy mentioned above is tested in our simulator, as we adopt a CFA approach by introducing parametric modifications in power capacities constraints and run a DLA to schedule the operation of the next hour. We also use a PFA, with simple analytical functions that define an order of priority of sources to cover demand in real-time operation.

## 4 | POLICY 1

The first policy we propose simulates a planning tool for a grid operator. To plan the next hour, we consider that a DLA would be run once an hour with 15 minute increments, which is short enough to capture the influence of ramping constraints (Section 4.1). A PFA is applied to decide the real-time operation by compensating wind generation forecast errors with adjustments in hydro output (Section 4.2).

We collectively represent the policy 1 as:

$$X_t^{\pi 1}(S_t) = \left( X_t^{\pi 1-DLA}(S_t), X_t^{\pi 1-PFA}(S_t) \right). \tag{36}$$

### 4.1 | Policy 1 lookahead

Understanding the time-steps is very important to understand this policy. The lookahead is run once an hour with 15 minute increments, which means that it is run every time $t \in \mathcal{T}$, where $\mathcal{T} = \{0, 4, 8 \ldots T\}$. The lookahead horizon is 24 hours ($H$ = 96). Figure 3 shows the framework of the lookahead.

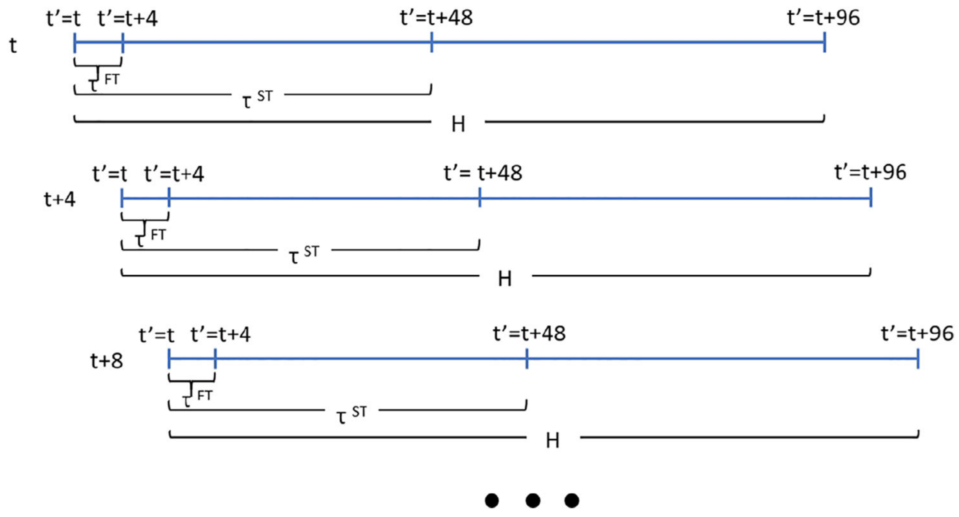


**FIGURE 3** Lookahead framework [Colour figure can be viewed at wileyonlinelibrary.com]

As commitments are made in advance here, we define the notification time $\tau$ as the number of 15 minute time-steps in the future for which we plan the generation of some sources. The notification times for all the energy sources in the system are:

- $\tau^{\mathrm{ST}} = 48$ (12 hours in advance for slow thermal plants)
- $\tau^{\mathrm{FT}} = 4$ (1 hour in advance for fast thermal plants)
- $\tau^{\mathrm{SD}} = 4$ (1 hour in advance for storage discharge)
- $\tau^{\mathrm{SC}} = 4$ (1 hour in advance for storage charge)
- $\tau^{\mathrm{EE}} = 4$ (1 hour in advance for energy exchange)

In the DLA, wind is treated deterministically and Equation (37) can be optimized.

$$X_t^{\pi 1-DLA}(S_t) = \underset{\tilde{x}_{t,t},\ldots,\tilde{x}_{t,t+H}}{\operatorname{argmin}} \sum_{t'=t}^{t+H} C\left(\tilde{S}_{tt'}\tilde{x}_{tt'}\right), \tag{37}$$

subject to Equations (6) to (19), as well as Equations (38) to (41).

To address the notification time requirement, we fix the decision variables from $t' = t$ to $t' = t + \tau$ with the decisions made before $t$. The following constraint has to be satisfied:

$$\{x_{tt'}\}_{t'=t}^{t+\tau-1} = \{x_{t-4,t'}\}_{t'=t+\tau}^{t+\tau+3}. \tag{38}$$

In this policy, we add another constraint to limit the use of hydro resources in the lookahead horizon $H$. The need for that constraint comes from the fact that $H$ is finite and relatively short (24 hours) considering that it takes months to fill some reservoirs, depending on its capacity and incoming water flow. This short time horizon might lead to an optimized solution that uses all hydro resources (or most of it) during the optimization horizon, but would actually leave the system without hydro resources for the following periods. For this reason, a total hydro generation resource $R_{t,t}^{\mathrm{HD}}$ available for the lookahead horizon is set. Hydro availability in each time-step evolves according to the transition function in Equation (40).

$$R_{t,t}^{HD} = R_0^{HD}, \tag{39}$$

$$R_{t,t'+1}^{HD} = R_{tt'}^{HD} - x_{tt'}^{HD} \geq 0. \tag{40}$$

To deal with the uncertainty related to wind forecast errors, we plan hydro reserves to cover demand when observed wind generation is lower than forecast. For that purpose, we adopt a CFA approach to introduce a parametric modification in Equation (7) for the lookahead, where a tunable parameter $\theta^{\mathrm{HD}} \in [0,\ 1]$ multiplies the hydro maximum power capacity for the next hour:

$$\left\{\tilde{x}_{tt'}^{HD}\right\}_{t'=t+4}^{t+7} \leq P_{max}^{HD}\left(1-\theta^{HD}\right). \tag{41}$$

The parameter $\theta^{\mathrm{HD}}$ is tuned with a one dimensional search, with a 0.05 step-size, as detailed in Section 8.

The results from the DLA are planned hydro generation, fast thermal generation, storage discharge, storage charge, and energy exchange for the next hour, and planned slow thermal generation for the period between 12 and 13 hours ahead. These results are input data for the real-time operation policy, a PFA described below.

## 4.2 | Policy 1 PFA

A PFA is used to define hydro generation in real-time operation. This adjustment in hydro is necessary to address the difference between forecast and observed wind generation. In reality, some thermals have the ability to vary their outputs with short notice. However, in the Brazilian grid, most of this adjustment is made by hydro plants and, therefore, we reduce the dimensionality of the problem allowing only hydro to provide real-time reserve without significantly impacting the results.

When hydro cannot meet demand because of power or ramping constraints, or because reserve is not sufficient, we have a deficit. The following equation represents our PFA rule for real-time operation:

$$X^{\pi 1-\mathrm{PFA}} = \begin{pmatrix} \left(x_t^{ST}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{ST},t'}^{ST}\right)_{t'=t+\tau^{ST}}^{t+\tau^{ST}+3}, \forall t \in \mathcal{T} \\ \left(x_t^{FT}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{FT},t'}^{FT}\right)_{t'=t+\tau^{FT}}^{t+\tau^{FT}+3}, \forall t \in \mathcal{T} \\ \left(x_t^{SD}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{SD},t'}^{SD}\right)_{t'=t+\tau^{SD}}^{t+\tau^{SD}+3}, \forall t \in \mathcal{T} \\ \left(x_t^{SC}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{SC},t'}^{SC}\right)_{t'=t+\tau^{SC}}^{t+\tau^{SC}+3}, \forall t \in \mathcal{T} \\ \left(x_t^{EE}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{EE},t'}^{EE}\right)_{t'=t+\tau^{EE}}^{t+\tau^{EE}+3}, \forall t \in \mathcal{T} \\ x_t^{HD} = max(P_{min}^{HD}, x_{t-1}^{HD} + \rho_{min}^{HD}, min(P_{max}^{HD}, x_{t-1}^{HD} + \rho_{max}^{HD}, \\ D_t - P_t^E - P_t^{PV} - x_t^{ST} - x_t^{FT} - x_t^{SD} + x_t^{SC} - x_t^{EE})) \\ x_t^{DF} = D_t - x_t^{HD} - P_t^E - P^{PV} - x_t^{ST} - x_t^{FT} - x_t^{SD} + x_t^{SC} - x_t^{EE}. \end{pmatrix} \tag{42}$$

Because we consider demand and solar generation forecasts perfect, $D_t = f_{tt'}^D$ and $P_t^{PV} = f_{tt'}^{PV}$. The generation from other sources, storage system charge, and energy exchange ($\tilde{x}_{tt'}^{ST}, \tilde{x}_{tt'}^{FT}, \tilde{x}_{tt'}^{SD}, \tilde{x}_{tt'}^{SC}, \tilde{x}_{tt'}^{EE}$) are previously decided by the DLA. Here, they belong to the state variable $S_t$. $E_t$ is the observed wind generation vector, which is different from $f_{tt'}^E$ for $t' > t$ and, therefore, require real-time adjustments from hydro.

## 5 | POLICY 2

The only difference of this policy from policy 1 is that, here, reserves are not provided by hydro anymore, but only by the storage system. This difference between policies 1 and 2 may seem subtle; however, not only does it change the source of spinning reserve but it makes the storage system compromise between two services. By adding a tunable parameter $\theta^S$ that sets a minimum stored energy in the system, it can still provide time-shifting, but its capacity decreases as $\theta^S$ increases.

### 5.1 | Policy 2 lookahead

In this policy, the lookahead is almost the same from the policy 1, except for the substitution of the constraint in Equation (41) by a new one that, instead of modifying hydro capacity, sets a minimum amount of energy in the storage system to provide real-time reserves (Equation (43)):

$$\left(R_{tt'}^S\right)_{t'=t+4}^{t+7} \geq \theta^S R_{\max}^S. \tag{43}$$

The parameter $\theta^S$ is tuned with a one dimensional search, with a 0.05 step-size, as detailed in Section 8.

Because hydro does not provide reserves in this policy, we also set a 1-hour notification time for this source ($\tau_t^{SC} = 4$).

### 5.2 | Policy 2 policy PFA

In this policy, as opposed to the policy 1, the operation of the storage system is not fixed as previously decided in the lookahead, which means that $\tau_t^{SD} = \tau_t^{SC} = 0$. Here, instead of hydro generation, storage discharge and charge are adjusted to fit wind forecast errors, while other sources of generation have already been fixed according to their notification time.

Let $B_t$ be, at time $t$, the difference between the demand and the sum of generation from all sources that were previously decided or observed, including the planned energy charge ($B_t = D_t - x_t^{\mathrm{HD}} - P_t^E - P_t^{\mathrm{PV}} - x_t^{\mathrm{ST}} - x_t^{\mathrm{FT}} + \tilde{x}_t^{\mathrm{SC}} - x_t^{\mathrm{EE}}$ ). In other words, $B_t$ is the remaining energy balance at time $t$. Here, when $B_t$ is positive, energy can be discharged from the storage system. Energy can also be charged to the storage system if $B_t$ is negative.

$$X^{\pi 2-PFA} = \left( \begin{array}{l} \left(x_t^{HD}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{HD},t'}^{HD}\right)_{t'=t+\tau^{HD}}^{t+\tau^{HD}+3} \\ \left(x_t^{ST}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{ST},t'}^{ST}\right)_{t'=t+\tau^{ST}}^{t+\tau^{ST}+3} \\ \left(x_t^{FT}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{FT},t'}^{FT}\right)_{t'=t+\tau^{FT}}^{t+\tau^{FT}+3} \\ \left(\tilde{x}_t^{SC}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{SC},t'}^{SC}\right)_{t'=t+\tau^{SC}}^{t+\tau^{SC}+3} \\ \left(x_t^{EE}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{EE},t'}^{EE}\right)_{t'=t+\tau^{EE}}^{t+\tau^{EE}+3} \\ B_t = D_t - x_t^{\mathrm{HD}} - P_t^E - P_t^{PV} - x_t^{ST} - x_t^{FT} + \tilde{x}_t^{SC} - x_t^{EE} \\ B_t \geq 0 \left( \begin{array}{l} x_t^{SC} = max\left(0, \tilde{x}_t^{SC} - B_t\right) \\ x_t^{SC} > 0 \left( \begin{array}{l} x_t^{SD} = 0 \\ x_t^{DF} = 0 \end{array} \right) \\ x_t^{SC} = 0 \left( \begin{array}{l} x_t^{SD} = min\left(R_t^S, x_{t-1}^{SD} + \rho_{max}^{SD}, P_{max}^{SD}, B_t - \tilde{x}_t^{SC}\right) \\ x_t^{DF} = B_t - \tilde{x}_t^{SC} - x_t^{SD} \end{array} \right) \end{array} \right) \\ B_t < 0 \left( \begin{array}{l} x_t^{SC} = min\left(\tilde{x}_t^{SC} - B_t, R_{max}^S - R_t^S, x_{t-1}^{SC} + \rho_{\max}^{SC}, P_{max}^{SC}\right) \\ x_t^{SD} = 0 \\ x_t^{DF} = 0 \end{array} \right) \end{array} \right) \tag{44}$$

## 6 | POLICY 3

This policy is a combination of policies 1 and 2. Here, both hydro and storage system can provide reserves. The storage system can still provide time-shifting but limited by reserve needs.

### 6.1 | Policy 3 lookahead

This policy combines the lookaheads of policies 1 and 2, with two constraints modified by tunable parameters. Here, the lookahead is subject to Equations (41) and (43):

The parameter $\theta^S$ is tuned together with $\theta^{\mathrm{HD}}$ in a grid search, as described in Section 8.

### 6.2 | Policy 3 policy PFA

In this policy, the operations of hydro and storage system are not fixed as previously decided in the lookahead, which means that $\tau_t^{\mathrm{HD}} = \tau_t^{\mathrm{SD}} = \tau_t^{\mathrm{SC}} = 0$. Hydro generation, storage discharge, and storage charge are adjusted to fit wind forecast errors, while other sources of generation have already been fixed according to their notification time.

We also make a small change in $B_t$, defined in policy 2, removing predecided hydro generation from the energy balance and introducing a planned hydro generation instead ($B_t = D_t - \tilde{x}_t^{HD} - P_t^E - P_t^{PV} - x_t^{ST} - x_t^{FT} + \tilde{x}_t^{SC} - x_t^{EE}$).

Because we have two sources that can provide reserves, we prioritize hydro to cover $B_t$ and, when it is not sufficient, energy can be discharged from the storage system. Energy can also be charged to the storage system if $B_t$ is negative.

$$X^{\pi 3-PFA} = \left( \begin{array}{l} \left(\tilde{x}_t^{HD}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{HD},t'}^{HD}\right)_{t'=t+\tau^{HD}}^{t+\tau^{HD}+3} \\ \left(x_t^{ST}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{ST},t'}^{ST}\right)_{t'=t+\tau^{ST}}^{t+\tau^{ST}+3} \\ \left(x_t^{FT}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{FT},t'}^{FT}\right)_{t'=t+\tau^{FT}}^{t+\tau^{FT}+3} \\ \left(\tilde{x}_t^{SC}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{SC},t'}^{SC}\right)_{t'=t+\tau^{SC}}^{t+\tau^{SC}+3} \\ \left(x_t^{EE}\right)_{t=t}^{t+3} = \left(\tilde{x}_{t-\tau^{EE},t'}^{EE}\right)_{t'=t+\tau^{EE}}^{t+\tau^{EE}+3} \\ B_t = D_t - \tilde{x}_t^{HD} - P_t^E - P_t^{PV} - x_t^{ST} - x_t^{FT} + \tilde{x}_t^{SC} - x_t^{EE} \\ B_t \geq 0 \left( \begin{array}{l} x_t^{HD} = \max\left(P_{min}^{HD}, x_{t-1}^{HD} + \rho_{min}^{HD}, min\left(P_{min}^{HD}, x_{t-1}^{HD} + \rho_{max}^{HD}, B_t + \tilde{x}_t^{HD}\right)\right) \\ B_t + \tilde{x}_t^{HD} - x_t^{HD} \geq 0 \left( \begin{array}{l} x_t^{SC} = max\left(0, \tilde{x}_t^{SC} - B_t - x_t^{HD}\right) \\ x_t^{SC} > 0 \left( \begin{array}{l} x_t^{\mathrm{SD}} = 0 \\ x_t^{DF} = 0 \end{array} \right) \\ x_t^{SC} = 0 \left( \begin{array}{l} x_t^{SD} = \min\left(R_t^S, x_{t-1}^{SD} + \rho_{max}^{SD}, P_{max}^{SD}, B_t - x_t^{HD} - \tilde{x}_t^{SC}\right) \\ x_t^{DF} = B_t - x_t^{HD} - \tilde{x}_t^{SC} - x_t^{SD} \end{array} \right) \end{array} \right) \\ B_t + \tilde{x}_t^{HD} - x_t^{HD} < 0 \left( \begin{array}{l} x_t^{SC} = \min\left(\tilde{x}_t^{SC} - B_t - x_t^{HD}, R_{max}^S - R_t^S, x_{t-1}^{\mathrm{SC}} + \rho_{\max}^{\mathrm{SC}}, P_{max}^{SC}\right) \\ x_t^{SD} = 0 \\ x_t^{DF} = 0 \end{array} \right) \end{array} \right) \\ B_t < 0 \left( \begin{array}{l} x_t^{HD} = \tilde{x}_t^{HD} \\ x_t^{SC} = min\left(-B_t + \tilde{x}_t^{SC}, R_{max}^S - R_t^S, x_{t-1}^{SC} + \rho_{max}^{SC}, P_{max}^{SC}\right) \\ x_t^{SD} = 0 \\ x_t^{DF} = 0. \end{array} \right) \end{array} \right) \tag{45}$$

## 7 | MODELING UNCERTAINTY

Three exogenous uncertain processes need to be modeled: forecasts for wind generation, solar generation, and energy demand. Demand is relatively predictable, especially for a large-grid like the Northeast subsystem in Brazil and, for that reason, its forecasts are considered perfect in the model, reducing the dimensionality of the problem. We also assume that forecasts for solar generation are perfect because, regardless of being hard to predict in days with spotted clouds, solar generation is still considerably smaller than other sources. It represents around 5% of demand in its peak of generation on a sunny day, thus its variability does not significantly affect the operation of the grid.

Energy exchange prices are based on spot market prices in Brazil and it could be uncertain if we were modeling an energy system in a country with more volatile prices or if the optimization horizon was longer. However, considering how spot market prices are formed in Brazil, energy exchange prices can be also treated as deterministic in the model.

Spot market prices in Brazil are defined on a weekly basis for each subsystem and for three different demand levels: high (from 6:00 PM to 9:00 PM), medium (from 7:00 AM to 6:00 PM and from 9:00 PM to 0:00 AM) and low (from 0:00 AM to 7:00 AM). These periods are slightly altered during daylight saving times.[17] This price arrangement by demand levels, however, is in a process of change and, by 2020, hourly prices will be defined on a daily basis for the next day. Even with this new methodology, spot market prices for the next day will be known in advance.

### 7.1 | Wind generation

Wind speed forecasts were modeled using the same technique presented in Ghadimi et al .[45] This method allows us to control the error in the forecast in our simulations. To create a realistic nonstationary behavior of wind forecasts, we add synthetic errors to the forecast wind speeds available at time $t$ to create another forecast series available at time $t + 4$ (1 hour later). Letting $f^V_{tt'}$ be the forecast of wind speed at time $t$ for time $t'$, $V_t$ be observed wind speed at time $t$ and assuming $\{ f^V_{0,t'} \}^H_{t'=0} = \{V_t\}^H_{t=0}$, we define:

$$f^V_{t+4,t'} = f^V_{t,t'} + \epsilon_{t+4,t'}, \tag{46}$$

where $\epsilon_{t+4,t'}$ is the noise.

To create this noise, a symmetric matrix $\sum$ is constructed such that $\sum(i,j) = \sigma^2_V e^{-\alpha|i-j|}$, where the values of $\sigma^2_V$ and $\alpha$ are constant and defined by the user to manipulate the quality of the forecasts. Then, a noise vector is defined as $\epsilon_t = C \times Z_t$, where $C$ is the lower triangular Cholesky decomposition of $\sum$ and $Z_t \tilde{\mathcal{N}}(0, I_{HxH})$ . For more details, refer to Reference 45.

The wind speed forecasts $f^V_{tt'}$ were generated on an hourly basis. However, to capture ramping effects, we used Brownian bridge interpolation to create data discretized in 15 minute intervals.

Wind speed forecasts $f^V_{tt'}$ were converted into wind power forecasts $f^E_{tt'}$ with a speed to power curve (Figure 4) developed by Reference 46.

### 7.2 | Solar generation

Except for days with spotted clouds, solar generation follows well-established patterns whose values depend on the cloudiness. We could consider three different patterns for the simulations, which would be for sunny, cloudy, and with spotted cloudy days. The pattern for cloudy days would follow the same pattern for sunny days but with lower values. The pattern for days with spotted clouds would have equal or lower values than the one for sunny days. Nevertheless, because solar generation in the modeled subsystem is too small, it does not considerably affect the simulation results and, therefore, only one pattern for sunny days will be considered for solar generation forecast in the article.

We recognize that, if solar capacity grows significantly in Brazil, its variability may affect the planning of the operation. However, the uncertainty modeled in wind forecasts is sufficient to test the abilities of the model in dealing with stochastic processes.

### 7.3 | Energy demand

Basically, the two main aspects that affect daily energy demand profiles are temperature and whether it is a business day or not. However, the region modeled in this work, Brazilian Northeast, has almost constant warm temperatures during all year. We can infer then that only business and nonbusiness days affect its energy demand profiles, which makes them reasonably predictable. For that reason, demand forecasts were modeled as perfect. We simply consider that demand forecasts are equal observed data: $f^D_t = D_t$.

The Brazilian grid operator provides demand data for the Northeastern subsystem discretized in 1 hour periods. To generate 15 minute time-step series, a simple linear interpolation is used.

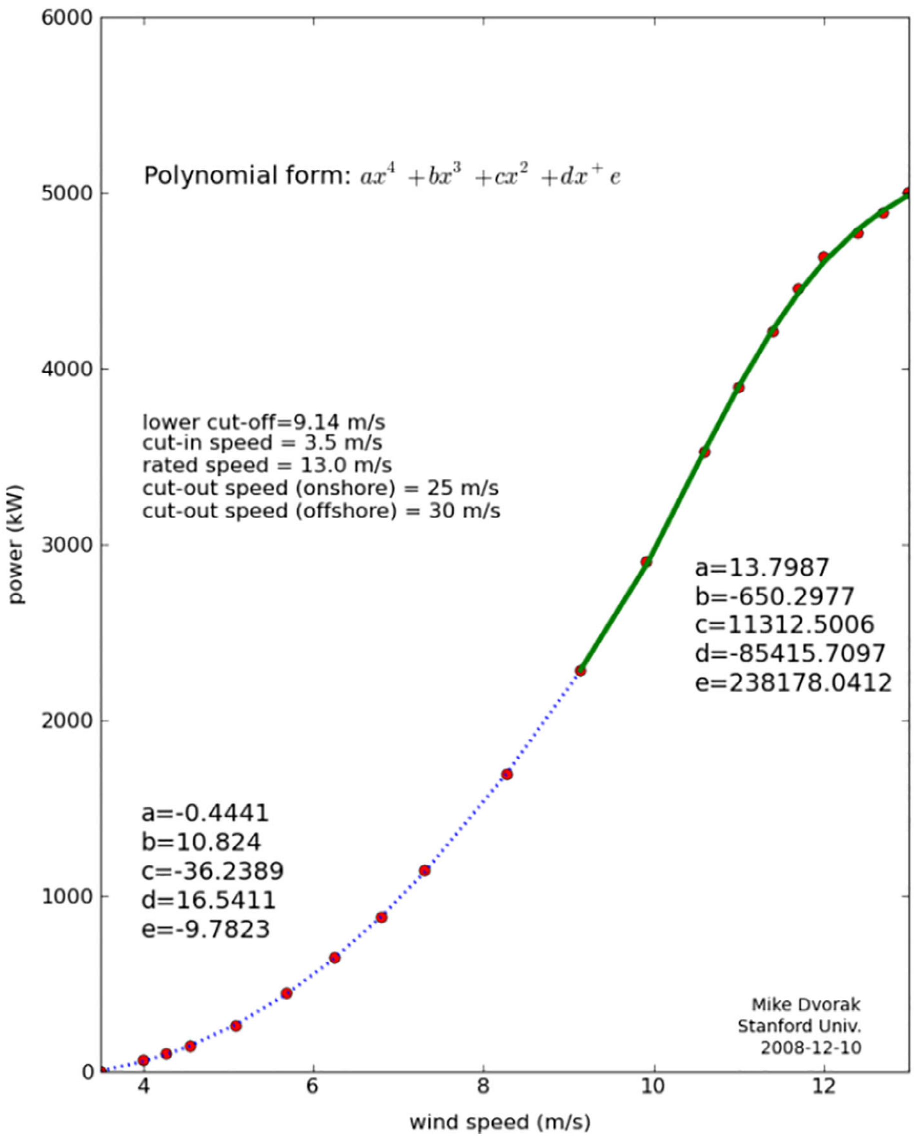


**FIGURE 4** Wind speed to power curve[46] [Colour figure can be viewed at wileyonlinelibrary.com]

## 8 | POLICY SEARCH

The policies 1, 2, and 3 are defined, but we still need to find the values of the parameters $\theta$ that give the best performance for each policy. We do that with policy search, which aims to find values of $\theta$ for which the policy $X^{\pi}(S_t| \theta)$ performs well. Equation (47) represents the policy search:

$$\min_{\theta \in \Theta} \mathbb{E}\left\{ \sum_{t=0}^{T} C(S_t, X_t^{\pi}(S_t, \theta)|S_0) \right\}. \tag{47}$$

where $S_{t+1} = S^M(S_t, x_t, W_{t+1})$ and $(S_0, W_1, ..., W_t...)$ is a given stochastic process.

The higher the dimension of $\theta$, the bigger the challenge of finding its optimal values, as randomly trying different values of $\theta$ can be highly inefficient.[47] For those cases, $\theta$ can be tuned using classic stochastic optimization algorithms. In our case, however, $\theta$ in policies 1 and 2 have only one dimension and, in policy 3, two dimensions. This means that we can do a full grid search for these cases.

While the dimensions of $\theta$ are not a problem here, the stochasticity of the simulations requires some attention. As explained in Section 7.1, wind forecasts for time $t + 4$ were generated by capturing the errors in the forecasts for time $t$. The SD of the forecast errors were set at 10% of the average wind speed at time $t$ for the lookahead horizon ($\sigma = 0.10 \sum_{t'=t}^{t'=t+H} f_{t,t'}^{V} / H$). To handle that noise, we run a thousand simulations for each $\theta$, and the $\theta$ which gives the lowest average cost over the simulations is the optimal.

We remind the reader that the policies are tuned offline, which means that the computational time to tune the parameters is not relevant to the online electricity dispatch because it is only done once. After the parameters are tuned, the optimization processes used to make the decisions in every time-step are simple and are calculated almost instantaneously.

The proposed policies were tuned for four different sets of input data. The datasets are differentiated on the basis of the capacities of the hydro and storage system, as presented in Table 3.

All the other deterministic data used in the tuning process, such as power capacities and ramping capacities, are real data from the Brazilian Northeast subsystem.

## 8.1 | Tuning policy 1

Policy 1 has only one tunable parameter, $\theta^{\mathrm{HD}}$. This parameter was tuned with a simple one dimensional search, in which $\theta^{\mathrm{HD}}$ ranged from 0 to 1 with a 0.05 step-size that looks for the lowest average costs of the problem over a thousand simulations solved with policy 1. Figure 5 shows the results of each search for different datasets:

**TABLE 3** Datasets for policy search

| Dataset | Hydro power capacity (GW) | Storage system capacity (GWh) |
|---|---|---|
| 1 | 2.0 | 1.5 |
| 2 | 2.0 | 3.0 |
| 3 | 4.0 | 1.5 |
| 4 | 4.0 | 3.0 |

With this policy, the storage system can only provide time-shifting. When we compare dataset 1 to 2, and dataset 3 to 4, we can see that increasing storage capacity does not significantly reduces costs. This shows that the storage system cannot reduce costs only by providing time-shifting.

Comparing dataset 1 to 3, and dataset 2 to 4, we conclude that there is an optimal hydro reserve that changes as a percentage of total hydro capacity, but does not change its absolute value. For the simulated grid, the optimal hydro reserve is 1.4 GW, which means $\theta^{\mathrm{HD}} = 0.70$ for datasets 1 and 2, and $\theta^{\mathrm{HD}} = 0.35$ for datasets 3 and 4.

## 8.2 | Tuning policy 2

Just as with policy 1, policy 2 also has only one tunable parameter, $\theta^{S}$, which was also tuned with a one dimensional search, in which $\theta^{S}$ ranged from 0 to 1 with a 0.05 step-size. Figure 6 shows the results of each search for different datasets:

With policy 2, the storage system can provide both reserves and time-shifting. Results from tuning this policy show that, after a certain value, costs does not significantly change when we increase $\theta^{S}$. We can conclude from this that there is an optimal value of storage

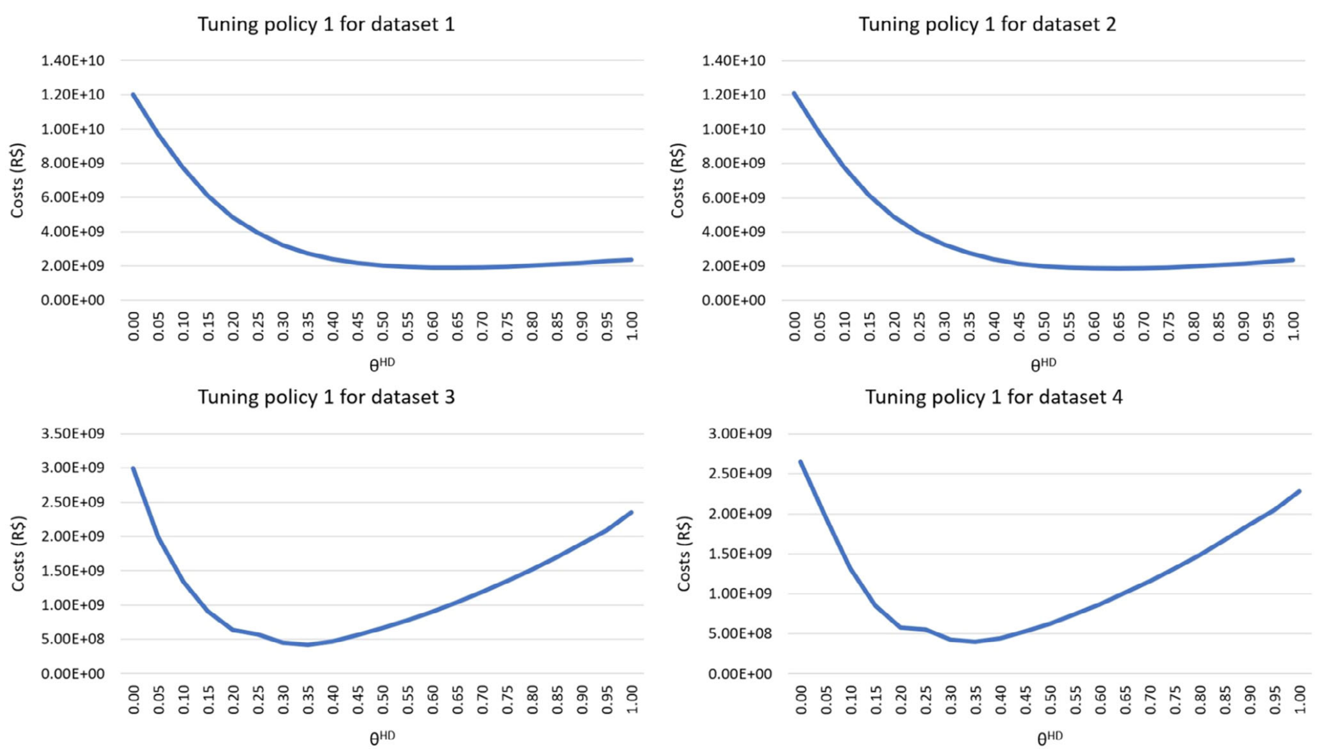


**FIGURE 5** Parameter tuning for policy 1 [Colour figure can be viewed at wileyonlinelibrary.com]

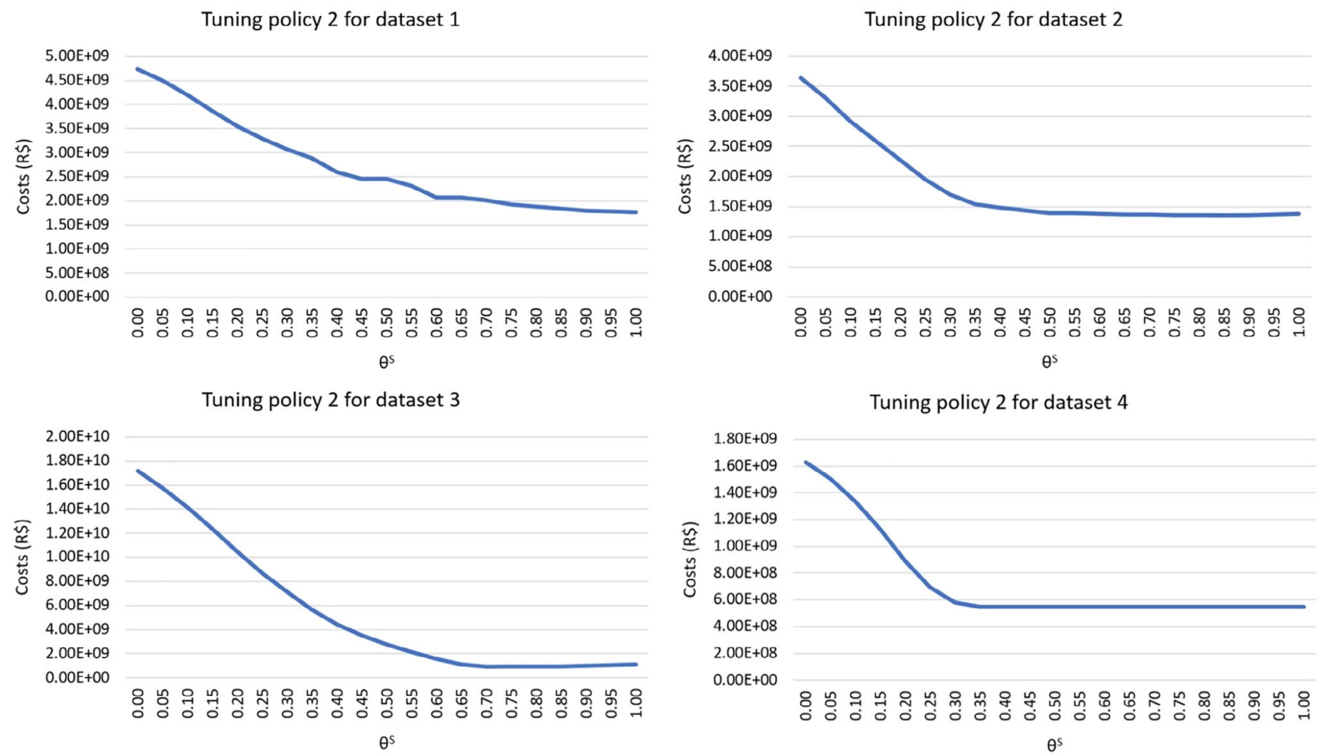


**FIGURE 6** Parameter tuning for policy 2 [Colour figure can be viewed at wileyonlinelibrary.com]

reserves and any capacity larger than that does not help reducing costs.

As $\theta^S$ gets bigger, time-shifting capacity is reduced. However, costs do not increase when we reduce the time-shifting capability of the storage systems. This confirms what was concluded in the results from tuning policy 1: for the simulated grid, time-shifting does not help reducing costs significantly.

### 8.3 | Tuning policy 3

Policy 3 has two tunable parameters, $\theta^{HD}$ and $\theta^S$. These parameters were tuned with a grid search, in which each $\theta$ ranged from 0 to 1 with a 0.05 step-size, to find the lowest average costs over a thousand simulations solved with policy 3. Figure 7 shows the results of each grid search for different datasets:

Results from tuning policy 3 show that, usually, storage systems should be prioritized to provide reserves over hydro, as $\theta^S$ is usually greater than $\theta^{HD}$. This can be explained by the fact that hydro reserves have an indirect potential cost, as reserved power may not be used in real-time operation. In that case, hydro (the cheapest source of energy in the system) is not being used in its total capacity and another source (a more expensive one) is scheduled to cover that difference.

## 9 | PERFORMANCE ANALYSIS

To evaluate the performance of each of the three proposed policies, we use as a benchmark a fourth policy in which we solve a deterministic nonlinear program for the whole simulation horizon using actual observed wind generation. This benchmark policy will produce an optimistic estimate of performance, because it is being allowed to see into the future.

Table 4 compares the results of each proposed policy with the know-in-advance policy. The performances $\Delta F^\pi(\theta)$ are calculated according to Equation (48). Because we are minimizing costs, smaller values for $\Delta F^\pi(\theta)$ mean better performances.

$$\text{Performance} = \Delta F^\pi(\theta) = \frac{F^\pi(\theta)}{F^{KA}}, \tag{48}$$

where $F^\pi(\theta)$ is the total cost over the simulation horizon with policy $\pi$ and $F^{KA}$ is the total cost with the know-in-advance policy.

The optimal values of $\theta$ varied according to the input dataset, but in all of them policy 2 outperformed policy 1. This shows that, for the modeled grid, it is economically advantageous to use storage system for spinning reserves while scheduling generation for 100% of hydro capacities (leaving no hydro reserves), than using hydro

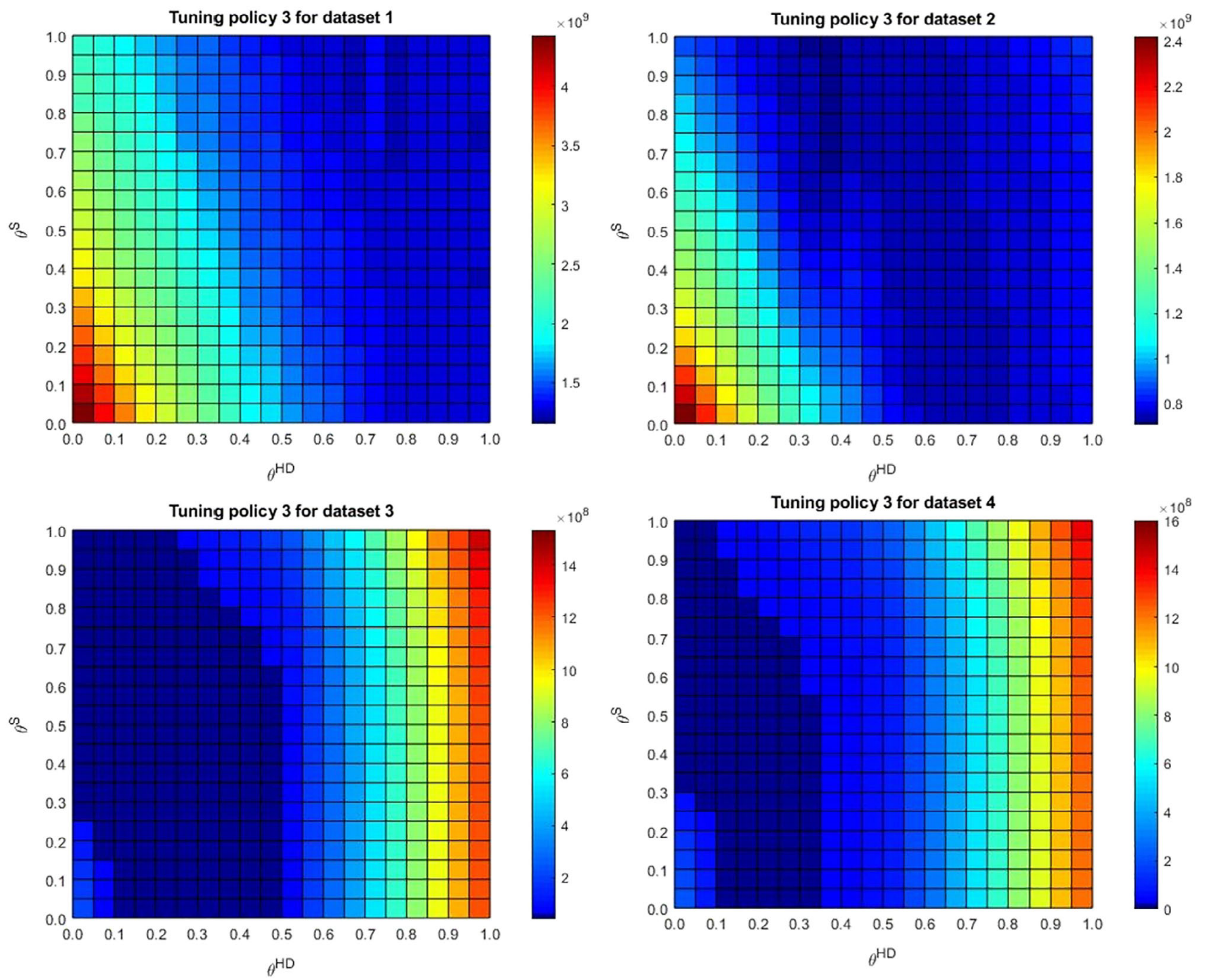


**FIGURE 7** Parameter tuning for policy 3 [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 4** Optimal tunable parameters and performances of each policy for different input datasets

| Dataset | Hydro power capacity (GW) | Storage system capacity (GWh) | Policy 1 | | Policy 2 | | Policy 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | $\theta^{HD}$ | $\Delta F^{\pi 1}$ | $\theta^S$ | $\Delta F^{\pi 2}$ | $\theta^{HD}$ | $\theta^S$ | $\Delta F^{\pi 3}$ |
| 1 | 2.0 | 1.5 | 0.70 | 168% | 1.00 | 121% | 1.00 | 0.70 | 117% |
| 2 | 2.0 | 3.0 | 0.70 | 210% | 0.85 | 118% | 0.35 | 1.00 | 109% |
| 3 | 4.0 | 1.5 | 0.35 | 122% | 0.70 | 116% | 0.00 | 0.70 | 111% |
| 4 | 4.0 | 3.0 | 0.35 | 173% | 0.35 | 100% | 0.00 | 0.35 | 100% |

for reserves while leaving storage system just for time-shifting.

Policy 3 outperformed both policies 1 and 2 for three datasets (and matched policy 2 for the fourth dataset), which shows that allowing hydro to provide reserves and storage systems to provide both time-shifting and reserves can be even more economical if reserves ($\theta^{HD}$ and $\theta^S$) are properly tuned.

In dataset 4, hydro and storage system capacities are enough for policies 2 and 3 to cover demand, matching the performance of the know-in-advance policy. In these simulations, after proper tuning of $\theta$, there was no deficit or energy imports. Thermal generation was only enough to satisfy the minimum output constraints and demand was supplied by intermittent sources, hydro and storage system.

## 10 | CONCLUSION

The article proposes an approach to design policies for energy dispatch with uncertain forecasts that uses a DLA policy with parametric modifications for scheduling and a simple PFA for real-time operation. The parameters are tuned in a stochastic simulator. These policies perform well when handling storage systems with stochastic forecasts and are easy to implement, but need to be properly tuned.

The lookahead with parametric modifications proposed here is a novel technique to deal with storage systems. It takes a lot of work to design the policies and test them, but once they are tuned, decisions can be computed almost instantaneously, which make the policies fast enough to be used in real-time operation of energy storage systems.

The strategy of setting reserves is widely used to deal with stochastic forecasts of generation or demand, but we show that tuning the reserves is important. In the simulations of this study, the optimal values of the tunable parameters for each proposed policy changed depending on the available hydro power and storage capacity, which indicates that they capture the characteristics of the modeled system and alter the optimal reserves depending on available hydro and storage capacities.

Policy 3, which allows both hydro and storage system to provide reserves, prioritizes the storage system to provide this service, leaving as much hydro power capacity as possible, which is the cheapest source in the system, available to be scheduled in the lookahead. Policy 3, when properly tuned, determines an optimal balance between the reserves of hydro and storage systems.

Policy 2 had better results than policy 1 because of the modeled grid which captured the characteristics of the Brazilian subsystem. However, policy 1 could have had better results in other grids. For instance, in countries with less fast thermal capacities and more variable spot market prices (represented here by energy imports), time-shifting could be more useful, which would favor policy 1, since this policy leaves all storage capacity for that service while hydro is the only source that provides spinning-reserves. In the grid modeled in this work, as we can see in Figure 6, storage capacity should be enough to provide reserves and any extra capacity for time-shifting does not significantly reduce costs.

Future research may include the application of parameterized lookaheads for type selection and sizing of storage systems based on grid needs for reserves and time-shifting that can be quantified with parameter tuning. The policies that were proposed here can also be applied, with only a few modifications, to grids without storage systems. For instance, spinning-reserves for thermal and hydro plants can be optimized with policy search.

## ACKNOWLEDGEMENTS

The authors would like to thank CAPES—Brazilian Federal Agency for Support and Evaluation of Graduate Education—for funding this work.

## ORCID

*Vinícius de Carvalho Neiva Pinheiro* https://orcid.org/0000-0002-4049-7256

## REFERENCES

1. Kaldellis J, Zafirakis D. Present situation and future prospects of electricity generation in aegean archipelago islands. *Energy Policy*. 2007;35(9):4623-4639. https://doi.org/10.1016/j.enpol.2007.04.004, https://www.sciencedirect.com/science/article/pii/S0301421507001425.
2. REN21, Renewables 2019 - global status report, Tech. rep; 2019.
3. Khosravi A, Machado L, Nunes R. Time-series prediction of wind speed using machine learning algorithms: a case study Osorio wind farm, Brazil. *Appl Energy*. 2018;224:550-566.
4. Operador Nacional do Sistema - ONS Histórico da operação, Tech. rep; 2019.
5. Huber M, Dimkova D, Hamacher T. Integration of wind and solar power in europe: assessment of flexibility requirements. *Energy*. 2014;69:236-246.
6. Zhou W, Lou C, Li Z, Lu L, Yang H. Current status of research on optimum sizing of stand-alone hybrid solar–wind power generation systems. *Appl Energy*. 2010;87(2):380-389.
7. Kaabeche A, Belhamel M, Ibtiouen R. Sizing optimization of grid-independent hybrid photovoltaic/wind power generation system. *Energy*. 2011;36(2):1214-1222.
8. Pinheiro V. Contribuição aos estudos regulatórios para inserição de sistemas de geração de energia elétrica compostos por fontes hidráulicas, solares e eólicas no brasil (Master thesis). College of Civil Engineering, Architecture and Urban Planning - State University of Campinas; 2016.
9. Blaga R, Sabadus A, Stefu N, Dughir C, Paulescu M, Badescu V. A current perspective on the accuracy of incoming solar energy forecasting. *Progr Energy Combust Sci*. 2019;70:119-144.
10. Miranda R, Ferreira PM, Schaeffer R, Szklod A. Limitations of thermal power plants to solar and wind development in Brazil; 2016.
11. Miranda R, Soria R, Schaeffer R, Szklo A, Saporta L. Contributions to the analysis of integrating large scale wind power into the electricity grid in the northeast of Brazil. *Energy*. 2017;118:1198-1209.
12. Lamadrid AJ, Mount T. Ancillary services in systems with high penetrations of renewable energy sources, the case of ramping. *Energy Econ*. 2012;34(6):1959-1971.
13. Keller V, English J, Fernandez J, et al. Electrification of road transportation with utility controlled charging: a case study for British Columbia with a 93% renewable electricity target. *Appl Energy*. 2019;253:113536.
14. Konzen G. *Planning for a new energy era: State of The Art of Energy Storage and Insertion of Intermittent Renewable Sources*, São Paulo: Empresa de Pesquisa Energética.
15. Nag PK. *Power Plant Engineering*. New Delhi: Tata McGraw-Hill Education; 2002.

16. Lacchini C, Dos Santos JCV. Photovoltaic energy generation in Brazil–cost analysis using coal-fired power plants as comparison. *Renew Energy*. 2013;52:183-189.
17. Libanori GHD, Pinheiro VDCN, Francato AL. Francato, Hybrid power plants as an alternative to fit pumped-storage hydro in brazilian electricity sector regulatory framework. *Int J Energy Res*. 2018;42:4898–4908. https://doi.org/10.1002/er.4224, https://onlinelibrary.wiley.com/doi/abs/10.1002/er.4224.
18. Ministério de Minas e Energia. *Empresa de Pesquisa Energética. Plano Decenal de Expansão de Energia 2026. Empresa de Pesquisa Energética*. Brasília: MME/EPE, 2017.
19. Castillo A, Gayme DF. Grid-scale energy storage applications in renewable energy integration: a survey. *Energ Conver Manage*. 2014;87:885-894.
20. Zhao H, Wu Q, Hu S, Xu H, Rasmussen CN. Review of energy storage system for wind power integration support. *Appl Energy*. 2015;137:545-553.
21. Keck F, Lenzen M, Vassallo A, Li M. The impact of battery energy storage for renewable energy power grids in Australia. *Energy*. 2019;173:647-657.
22. Wu W, Lin B. Application value of energy storage in power grid: a special case of China electricity market. *Energy*. 2018;165:1191-1199.
23. Zame KK, Brehm CA, Nitica AT, Richard CL, Schweitzer GD III. Smart grid and energy storage: policy recommendations. *Renew Sustain Energy Rev*. 2018;82:1646-1654.
24. Zhao B, Zhang X, Chen J, Wang C, Guo L. Operation optimization of standalone microgrids considering lifetime characteristics of battery energy storage system. *IEEE Trans Sustain Energy*. 2013;4(4):934-943.
25. Hu W, Chen Z, Bak-Jensen B. Optimal operation strategy of battery energy storage system to real-time electricity price in Denmark. Paper presented at: Proceedings of the IEEE PES General Meeting; 2010:1-7; IEEE.
26. Akhavan-Hejazi H, Mohsenian-Rad H. Optimal operation of independent storage systems in energy and reserve markets with high wind penetration. *IEEE Trans Smart Grid*. 2013;5(2):1088-1097.
27. Lo CH, Anderson MD. Economic dispatch and optimal sizing of battery energy storage systems in utility load-leveling operations. *IEEE Trans Energy Convers*. 1999;14(3):824-829.
28. Mercier P, Cherkaoui R, Oudalov A. Optimizing a battery energy storage system for frequency control application in an isolated power system. *IEEE Trans Power Syst*. 2009;24(3):1469-1477.
29. Appino RR, Ordiano JÁG, Mikut R, Faulwasser T, Hagenmeyer V. On the use of probabilistic forecasts in scheduling of renewable energy sources coupled to storages. *Appl Energy*. 2018;210:1207-1218.
30. Arnold M, Andersson G. Model predictive control of energy storage including uncertain forecasts. Paper presented at: Proceedings of the Power Systems Computation Conference (PSCC); Vol. 23, 2011:; Citeseer, Stockholm, Sweden.
31. Cheng B, Powell WB. Co-optimizing battery storage for the frequency regulation and energy arbitrage using multi-scale dynamic programming. *IEEE Trans Smart Grid*. 2016;9(3):1997-2005.
32. Varkani AK, Daraeepour A, Monsef H. A new self-scheduling strategy for integrated operation of wind and pumped-storage power plants in power markets. *Appl Energy*. 2011;88(12):5002-5012.
33. Drury E, Denholm P, Sioshansi R. The value of compressed air energy storage in energy and reserve markets. *Energy*. 2011;36(8):4959-4973.
34. Kazemi M, Zareipour H, Amjady N, Rosehart WD, Ehsan M. Operation scheduling of battery storage systems in joint energy and ancillary services markets. *IEEE Trans Sustain Energy*. 2017;8(4):1726-1735.
35. Archer C, Simão H, Kempton W, Powell WB, Dvorak M. The challenge of integrating offshore wind power in the us electric grid. Part i: wind forecast error. *Renew Energy*. 2017;103:346-360.
36. Simao H, Powell WB, Archer C, Kempton W. The challenge of integrating offshore wind power in the us electric grid. Part ii: simulation of electricity market operations. *Renew Energy*. 2017;103:418-431.
37. Lu Y, Zhao C, Watson J.-P, Pan K, Guan Y. Two-stage and multi-stage stochastic unit commitment under wind generation uncertainty. Paper presented at: Proceedings of the IIE Annual ConferenceInstitute of Industrial and Systems Engineers (IISE); 2014:4148.
38. Dai T, Qiao W. Trading wind power in a competitive electricity market using stochastic programing and game theory. *IEEE Trans Sustain Energy*. 2013;4(3):805-815.
39. Mesbah A. Stochastic model predictive control: an overview and perspectives for future research. *IEEE Control Syst Mag*. 2016;36(6):30-44.
40. Veldman E, Gibescu M, Slootweg J, Kling W. Technical benefits of distributed storage and load management in distribution grids. Paper presented at: Proceedings of the 2009 IEEE Bucharest PowerTech; 2009:18; IEEE.
41. Powell WB. A unified framework for optimization under uncertainty, optimization challenges in complex. *Netw Risky Syst*. 2016;45-83. https://pubsonline.informs.org/doi/10.1287/educ.2016.0149.
42. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT press; 2018.
43. Dalamagkidis K, Kolokotsa D, Kalaitzakis K, Stavrakakis GS. Reinforcement learning for energy conservation and comfort in buildings. *Build Environ*. 2007;42(7):2686-2698.
44. Powell WB. A unified framework for stochastic optimization. *Europ J Operat Res*. 2019;275(3):795–821.
45. Ghadimi S, Perkins RT, Powell WB. Parametric simulation optimization for multistage stochastic programming; 2019.
46. Dvorak MJ, Corcoran BA, Ten Hoeve JE, McIntyre NG, Jacobson MZ. Supplemental Online Material: Us east coast offshore wind energy resources and their relationship to peak-time electricity demand; 2011.
47. Jiang Daniel R, Pham Thuy V, Powell Warren B, Salas Daniel F, Scott Warren R. A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work? *IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning*. 2014. https://ieeexplore.ieee.org/document/7010626/authors#authors.

**How to cite this article:** V de Carvalho Neiva Pinheiro, Francato AL, Powell WB. Reinforcement learning for electricity dispatch in grids with high intermittent generation and energy storage systems: A case study for the Brazilian grid. *Int J Energy Res*. 2020;44:8635–8653. https://doi.org/10.1002/er.5551