Optimal Day-Ahead Trading and Storage of Renewable Energies - An Approximate Dynamic Programming Approach

Nils Löhndorf*, Stefan Minner

Department of Business Administration
University of Vienna
Brünner Str. 72
1210 Wien, Austria

December 11, 2009

Abstract

A renewable power producer who trades on a day-ahead market sells electricity under supply and price uncertainty. Investments in energy storage mitigate the associated financial risks and allow for decoupling the timing of supply and delivery. This paper introduces a model of the optimal bidding strategy for a hybrid system of renewable power generation and energy storage. We formulate the problem as a continuous-state Markov decision process and present a solution based on approximate dynamic programming. We propose an algorithm that combines approximate policy iteration with *Least Squares Policy Evaluation* (LSPE) which is used to estimate the weights of a polynomial value function approximation. We find that the approximate policies produce significantly better results for the continuous state space than an optimal discrete policy obtained by linear programming. A numerical analysis of the response surface of rewards on model parameters reveals that supply uncertainty, imbalance costs and a negative correlation of market price and supplies are the main drivers for investments in energy storage. Supply and price autocorrelation, on the other hand, have a negative effect on the value of storage.

Keywords: Renewable Energies, Energy Storage, Day-Ahead Market, Optimal Bidding Strategy, Approximate Dynamic Programming

*corresponding author, email: nils.loehndorf@univie.ac.at, phone: +43 1 4277 37956

1 Introduction

Many European countries today subsidize investments in renewable energies by guaranteeing a fixed rate for each kilowatt hour of electricity fed into the grid. Since this rate decreases over the years, a producer of renewable power is going to begin selling electricity directly at the market as soon as subsidies yield lower profits than direct trading. This combination of renewable power production and trading has created new investment opportunities but also challenges for power generating companies.

One challenge is that prices in electricity markets are uncertain. A wholesale electricity market typically consists of a day-ahead and a real-time market. At the day-ahead market, producers (e.g., generating companies) place supply bids and consumers (e.g., electricity providers) place demand bids which mature at the following day. After the day-ahead market is closed, the system operator announces a uniform clearing price which depends on the cumulated bids of all market participants. Since market participants do not reveal their bidding decisions, bidding eventually takes place under price uncertainty.

Another challenge is that the bidding volume may not match the physical volume generated by the renewable power source. In that case, the difference is cleared at the real-time market, also known as the balancing market. If the total difference of demand and supply is positive, the system operator activates operational reserves, which increases the real-time price. If the total difference is negative, the system operator deactivates operational reserves, which decreases the real-time price. Some markets, such as NordPool in Scandinavia, even create asymmetric real-time prices, so that a negative imbalance always settles above and a positive imbalance always below the day-ahead price. A renewable power producer therefore has to account for the cost that arises when a bid does not match supply. Moreover, in a market with a significant share of solar or wind power, producers are going to use the same weather forecasts, which leads to a correlation of forecast errors and real-time price deviations. Bidding decisions of renewable power producers therefore take place under price and supply uncertainty.

Most literature on optimal bidding strategies in electricity markets deals with supplier interaction and the commitment of thermal units. See Wen and David (2000) for an overview. Renewable power producers, however, are often too small to exercise market power and influence other player's decisions. Only few authors deal with the optimal bidding strategy of renewable power producers. Bathurst et al. (2002) propose a stochastic model to minimize the expected

cost of balancing. The authors assume that real-time prices are known in advance and develop a stochastic supply model from historical data. Matevosyan and Söder (2006) propose a two-stage stochastic program to minimize the cost of balancing of a wind power producer. The authors use price scenarios and model wind power supplies as an autoregressive (ARMA) model.

An even greater challenge arises when a renewable power producer has the additional option to store electricity. Storage can then be used as a hedge against the costs of balancing and allows the producer to respond to market prices. In addition to price and supply uncertainty, an optimal bidding strategy has to account for future price and supply realizations as well as future bidding decisions. The ability of energy storage to take advantage of arbitrage opportunities is addressed in Graves et al. (1999), although not in combination with renewable energies. Optimal bidding strategies for hybrid systems of wind power generation and energy storage are addressed in Bathurst and Strbac (2003), Korpås and Holen (2006) and Brunetto and Tina (2007) for deterministic supply and price paths. Uncertainty is considered in Fleten and Kristoffersen (2007) and García-González et al. (2008) who model optimal bidding strategies as two-stage stochastic programs. However, there still exists a research gap for models that consider both, supply and price uncertainty, as well as making bidding decisions over time.

Note that renewable power sources, such as wind or solar, do not have the natural means to control their energy output, and for them storage capacity requires additional investments. The most common storage scheme is pumped-hydro storage, where water is pumped into an elevated reservoir to consume excess electricity and released to generate electricity when it is needed later on. See Schainker (2004) for an overview of different energy storage technologies.

We propose to model the optimal bidding problem of a renewable power producer with storage as a continuous-state Markov decision process (MDP). A feature of the MDP is that an optimal bidding strategy derived from solving the problem not only considers price and supply uncertainty, but also takes future states and decisions into account. The difficulty with this formulation is that generic solution algorithms for MDPs, such as value or policy iteration, are subject to the *curse of dimensionality*. These algorithm are therefore only capable of computing an optimal bidding strategy for a small number of discrete states and decisions with known transition matrix. However, the state and action space of the bidding problem is continuous and a discretization with increasing resolution is computationally intractable.

To circumvent the curse of dimensionality and solve the problem efficiently, we propose a solution based on *approximate dynamic programming* (ADP). The ADP strategy presented in

this paper goes back to a class of methods known as *temporal difference* (TD) learning (Sutton and Barto, 1998). These methods learn the value function of an MDP by controlling the decision process according to some policy and iteratively updating the value function estimate. We will focus on a variant known as *least squares policy evaluation* (LSPE) (Nedic and Bertsekas, 2003) which approximates the value function by a set of linearly independent basis functions, where the function weights are estimated by least squares methods. A key advantage of using LSPE for our problem is that this method can handle a continuous state space. In line with Lagoudakis and Parr (2003), we combine LSPE with policy iteration to approximate the value function and to find a near-optimal policy, i.e., we iteratively find a near-optimal bidding strategy. For policy evaluation and improvement, our ADP algorithm has access to a simulation model of the stochastic decision process.

As a benchmark, we compute the transition matrix for an equivalent discrete state MDP and use linear programming to determine an optimal solution. We then compare the LP policy against two ADP policies with different value function approximations. Additionally, to study the influence of model parameters on discounted rewards, we analyze the response surface with a regression model.

The paper is organized as follows: In Section 2 we describe our model as well as the stochastic supply and price processes. In Section 3 we present two solution methods for the problem. Section 4 reports our numerical results, and Section 5 gives a summary and outlook.

2 Model Formulation

2.1 Markov Decision Process

We assume that during each day the renewable power producer places a bid at the day-ahead market before observing the realization of price and supply. If the realized supply is below the volume of a bid, the producer first empties the storage and then purchases the remainder at the real-time market. If the realized supply exceeds the bidding volume, the producer first stores excess supply before using the real-time market to clear the imbalance.

The power producer is assumed to be price-taker and the bidding strategy does not affect the bidding behavior of other market participants – think of a small player such as an individual wind farm operator. Since the marginal value of wind power is zero, we assume that the producer places a fixed-volume bid for any realization of the uncertain price. Price and supply follow an

autoregressive stochastic process so that both observations contain information on price and supply of the following day. A bid that does not match supply is automatically balanced by the system operator. We assume that the price of positive reserve is always u times higher and the price of negative reserve always o^{-1} times lower than the day-ahead price. We aggregate reservoir, pump and generator capacity as storage capacity. Charges and discharges are subject to losses; the total loss is referred to as *round-trip efficiency*.

Each period the power producer observes the current market price p as well as renewable power supplies y. The producer furthermore observes the amount of energy available in storage g at the end of period previous period. These three variables constitute a state of the process $S = \{y, p, g\} \in \mathcal{S}$, with $\mathcal{S}$ as the state space. The transition function $P(S'|x, S)$ denotes the probability that the next state will be S' given that the previous state was S and decision x was made. Based on this information, the producer makes a bidding decision $x \in \mathcal{X}$. Then, a random transition to the next state occurs, in which the bid matures, and the producer obtains a reward $r(S, x, S')$ that depends on the bid as well as the transition from S to S'.

The objective of the power producer is to select a policy $\pi(S)$ that assigns each state in S a decision in $\mathcal{X}$ such that the expected discounted cash flow of rewards is maximized. Let us state the objective function as the following infinite horizon, discounted Markov decision process,

$$V(S) = \max_{x \in \mathcal{X}} \left\{ \int_{S' \in \mathcal{S}} P(S'|S, x) \big(r(S, x, S') + \gamma V(S') \big) \, dS' \right\}, \tag{1}$$

with V being the value function and γ being the discount factor.

Denote C as storage capacity and c^+ (c^-) as the amount of energy charged (discharged) during a period and η^+ (η^-) as the efficiency of the charging (discharging) process with $0 < \eta^{\pm} \leq 1$, i.e.,

$$c^+ = \max\Big\{ \min\Big\{ y' - x, \frac{C - g}{\eta^+} \Big\}, 0 \Big\}, \tag{2}$$

$$c^- = \max\Big\{ \min\Big\{ x - y', \eta^- g \Big\}, 0 \Big\}. \tag{3}$$

The storage state transition from g to g' is deterministic for given realizations of price and supply. In case of a positive imbalance, the final storage level in the next period is C, and in case of a negative imbalance the final storage level is zero. The storage balance equation is given

by

$$g' = g + \eta^+ c^+ - \frac{c^-}{\eta^-}. \tag{4}$$

Rewards $r(S, x, S')$ are uncertain and depend on the capability to match the bid x with available capacity in the next period. If a bid exceeds renewable power supplies, the storage is discharged until $x - c^- - y' = 0$. Then the producer has to pay $u \cdot p'$ with $u > 1$ to the system operator for each unit of negative imbalance. If a bid is lower than renewable power supplies, the storage is charged until $y' - x - c^- = 0$. Then the producer receives $o \cdot p'$ with $0 \leq o < 1$ from the system operator for each unit of positive imbalance. The reward function is given by

$$r(S, x, S') = \begin{cases} (y' + c^-)p' - up'(x - y' - c^-) & \text{if } x > y', \\ xp' + op'(y' - x - c^+) & \text{otherwise.} \end{cases} \tag{5}$$

Note that the limitation of first using storage to clear an imbalance and then the real-time market may not be an optimal ex-post decision for all realizations of price and supply.

2.2 Stochastic Processes

In line with the literature, we assume that the stochastic processes of price and supply follow a first-order *autoregressive* process, i.e., an AR(1) process with normally distributed error terms. Moreover, as soon as multiple renewable power producers trade in the same market, their supplies will be positively correlated and the market price will move inversely proportional to the overall renewable supply – an effect that has already been observed with wind power supplies and spot market prices in Germany (Neubarth et al., 2006). We therefore additionally assume that the price is dependent on supply to account for the homogeneous trading patterns of renewable power producers.

We define mean, variance, autocorrelation and correlation of price and supply exogenously. Denote Y as the autoregressive process of supply with mean μ_Y, variance σ_Y^2 and autocorrelation $\theta_Y \in [0, 1)$. Since the realized supply y of the previous period partially explains the supply realization y of the current period, it has a direct effect on mean and variance of the overall stochastic process. The autoregressive supply process with given mean and variance is modeled

as

$$y' = \theta_Y y + \varepsilon^Y \quad \text{with} \quad \varepsilon^Y \sim N(\mu_Y^\varepsilon, \sigma_Y^\varepsilon), \tag{6}$$

$$\mu_Y^\varepsilon = (1-\theta_Y)\mu_Y, \tag{7}$$

$$\sigma_Y^\varepsilon = \sqrt{(1-\theta_Y^2)\sigma_Y^2}. \tag{8}$$

Equations (7) and (8) adapt the moments of the error term ε_t^Y, such that the supply process has mean μ_Y and variance σ_Y^2.

Denote P as the stochastic process of the price with mean μ_P, variance σ_P^2, autocorrelation $\theta_P \in [0,1)$ and parameter θ_{PY} to control the dependence of P on Y. In this case, the current price p' is partially explained by the realized price p of the previous period as well as the random change in supplies of the current period. The stochastic model is then given by

$$p' = \theta_P p + \theta_{PY}(y' - \theta_Y y) + \varepsilon^P, \quad \text{with } \varepsilon_t^P \sim N(\mu_P^\varepsilon, \sigma_P^\varepsilon), \tag{9}$$

$$\mu_P^\varepsilon = (1-\theta_P)\mu_P - \theta_{PY}(1-\theta_Y)\mu_Y, \tag{10}$$

$$\sigma_P^\varepsilon = \sqrt{(1-\theta_P^2)\sigma_P^2 - \theta_{PY}^2(1-\theta_Y^2)\sigma_Y^2}. \tag{11}$$

As before, equations (10) and (11) adapt the moments of the error term ε^P such that the price process has the intended mean and variance. Moreover, the parameter θ_{PY} controls the dependency of prices on supplies. Let us define the correlation of price and supply which is used as an exogenous parameter later on,

$$\rho = \frac{\theta_{PY}(1-\theta_Y^2)}{1-\theta_Y\theta_P}\sqrt{\frac{\sigma_Y^2}{\sigma_P^2}} \quad \text{with} \quad |\rho| \leq \frac{\sqrt{(1-\theta_Y^2)(1-\theta_P^2)}}{1-\theta_Y\theta_P}. \tag{12}$$

By rearranging terms, we can compute θ_{PY} as a function of autocorrelation and correlation of Y and P, which allows to model both processes with a given correlation. Note that Equation (11) is undefined for some θ_P, θ_Y and θ_{PY}, such that the correlation coefficient ρ is bounded from above and below.

3 Solution Methods

We approximate the optimal policy of the proposed Markov decision process by using an algorithm which iteratively combines policy evaluation and policy improvement. The algorithm

thereby approximates the value function of a given policy and then uses the learned relationship of state and reward to improve the policy.

To assess the solution quality of this approach, it is desirable to know the optimal policy. If we relax the assumption of a continuous state space and use a discrete state space instead, we can compute the transition matrix and have linear programming compute an optimal policy. We use this discrete policy as a benchmark for the ADP policy.

3.1 Computing the Transition Matrix

For a discrete solution of the value function (1), we need a discrete representation of the state transition function, which we refer to as transition matrix. Such a discrete state transition is characterized by discrete realizations of the random variables P and Y as well as a discrete storage state transition. As a result, we need a formulation of the conditional probabilities of price and supply defined over a discrete set of state values. To determine the conditional probability of supply $P_Y(y'|y)$, we restate (6), such that

$$\varepsilon^Y = y' - \theta_Y y, \tag{13}$$

With $\Phi_Y \sim N(\mu_Y^\varepsilon, \sigma_Y^\varepsilon)$ as the probability distribution of the random shock ε^Y, the cumulated probability of y' conditional on y then becomes

$$P_Y(y'|y) = \Phi_Y(y' - \theta_Y y). \tag{14}$$

If we define the stochastic supply process over the discrete set $Y^d = \{Y_L, ..., Y_U\}$ for the discrete probabilities P_Y^d it needs to hold that $\sum_{y' \in Y^d} P_Y^d(y'|y) = 1 \ \forall \ y$. To map the continuous distribution Φ_Y to the finite set Y^d, we round all real values to the next integer and cumulate the probability mass of the tails at the upper and lower bounds. Then, the truncated discrete conditional probability of supply is

$$P_Y^d(y'|y) = \begin{cases} \Phi_Y(Y_L - \theta_Y y + 0.5) & \text{if } y' = Y_L, \\ 1 - \Phi_Y(Y_U - \theta_Y y - 0.5) & \text{if } y' = Y_U, \\ \Phi_Y(y' - \theta_Y y + 0.5) - \Phi_Y(y' - \theta_Y y - 0.5) & \text{if } Y_L < y' < Y_U. \end{cases} \tag{15}$$

Accordingly, we derive the conditional probability of the price $P_P(p'|p, y', y)$ by mapping the distribution $\Phi_P \sim N(\mu_P^\varepsilon, \sigma_P^\varepsilon)$ to the discrete set $P^d = \{P_L, ..., P_U\}$. We restate (9) as before and the truncated discrete conditional probability of the price becomes

$$P_P^d(p'|p, y, y') = \begin{cases} \Phi_P(P_L - \theta_{Pp} - \theta_{PY}(y' - \theta_Y y) + 0.5) & \text{if } p' = P_L, \\ 1 - \Phi_P(P_U - \theta_{Pp} - \theta_{PY}(y' - \theta_Y y) - 0.5) & \text{if } p' = P_U, \\ \Phi_P(p' - \theta_{Pp} - \theta_{PY}(y' - \theta_Y y) + 0.5) & \\ \quad -\Phi_P(p' - \theta_{Pp} - \theta_{PY}(y' - \theta_Y y) - 0.5) & \text{if } P_L < p' < P_U. \end{cases} \tag{16}$$

To complete the discrete transition function, we restate the storage balance equation of (4). As it is difficult to model a discrete storage state transition with $\eta < 1.0$ we assume perfect round-trip efficiency so that the storage balance simplifies to

$$g' = \max\{\min\{y' + g - x, C\}, 0\}. \tag{17}$$

By this, we avoid additional rounding errors when using the discrete policy as benchmark to control the continuous-state process.

With this discretization of the stochastic process, we can formulate the probability transition matrix $\mathbb{P}(S'|S, x)$ which assigns a probability to each transition from a state S to a subsequent state S' after decision x has been made. The matrix is defined as

$$\mathbb{P}(S'|S, x) = \begin{cases} P_Y^d(y'|y) P_P^d(p'|y, y', p) & \text{if } g' = \max\{\min\{y' + g - x, C\}, 0\}, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

The transition probabilities are computed by multiplying the conditional probabilities of price and supply subject to a constraint on the storage balance. If (17) does not hold, the storage transition is infeasible and its probability set to zero.

3.2 Linear Programming Formulation

Denote $V(S)$ as the decision variable of the linear program. With $\mathbb{P}(S'|S, x)$ as the probability of a state transition from state S to state S' after decision x is made and $r(S, x, S')$ as the corresponding reward, we can state the discrete state, infinite horizon Markov decision process

as

$$\min_{V(S)\in\mathbb{R}} \sum_{s\in\mathcal{S}} \sum_{x\in\mathcal{X}} V(S) \tag{19}$$

$$s.t.\ V(S) \geq \sum_{S'\in\mathcal{S}} \mathbb{P}(S'|S,x)\big(r(S,x,S') + \gamma V(S')\big) \quad \forall \quad s\in\mathcal{S}, x\in\mathcal{X} \tag{20}$$

The optimal policy is found by selecting $\pi(S) = x$ where (20) is binding. In case no constraint is binding, an arbitrary decision is chosen (Puterman, 2005, p.223).

The tractability of this method is limited by the $|\mathcal{S}|$-dimensional decision vector with $|\mathcal{S}|\times|\mathcal{X}|$ inequality constraints and the necessity of a transition matrix. For the benchmark, we therefore compute the optimal policy only for a small discrete state space.

3.3 Least Squares Policy Evaluation

A widely used algorithm in approximate dynamic programming is temporal difference (TD) learning (Sutton and Barto, 1998) which combines Monte Carlo simulation with dynamic programming. Denote $\bar{V}$ as the post-decision value function which returns the expected value after a decision has been made,

$$\bar{V}(S,x) = \int_{S'\in\mathcal{S}} P(S'|S,x)\big(r(S,x,S') + \gamma V(S')\big)\ dS'. \tag{21}$$

Suppose that we use a simulation model to sample N state transitions, where each transition gives us a realization of the expectation. The TD learning algorithm updates the value estimate of a given state S and decision x by observing the reward and the discounted value of the subsequent state associated with the simulated state transition. The value estimate after n iterations is given by

$$\bar{V}^{n+1}(S^n,x^n) = \bar{V}^n(S^n,x^n) + \alpha^n\big(r(S^n,x^n,S^{n+1}) + \gamma\bar{V}^n(S^{n+1},x^{n+1}) - \bar{V}^n(S^n,x^n)\big), \tag{22}$$

where the parameter $\alpha^n \in (0,1]$ is used to smooth the updates of the estimates.

The updating step in (22) only works if we assume that the state space is discrete. Since we are dealing with a continuous state and action space, however, let us resort to another popular approximation architecture. Assume that $\bar{V}(S,x;w)$ is a linear combination of basis functions $\phi_i(S,x)$ with weights w_i and $k \in \{1,2,...,K\}$ which approximates the true value

function. Denote w and $\Phi(S,x)$ as the corresponding vectors of length K with $w^\top$ and $\Phi^\top$ as their transposes. Then, the approximate value function is defined as

$$\bar{V}(S,x;w) = \sum_{k=1}^{K} w_k \phi_k(S,x) = w^\top \Phi(S,x) \approx \bar{V}(S,x). \tag{23}$$

A basis function $\phi_k(S,x)$ may be any non-linear function of S and x. If all basis functions in $\Phi(S,x)$ are linearly independent, we can use ordinary least squares to estimate the weight vector w, such that

$$w = \left(\sum_{n=1}^{N-1} \Phi(S^n,x^n)\Phi^\top(S^n,x^n) \right)^{-1} \left(\sum_{N=1}^{N-1} \Phi(S^n,x^n) r(S^n,x^n,S^{n+1}) \right). \tag{24}$$

This gives us an approximation of the function of expected immediate rewards, which would be sufficient for $\gamma = 0.0$. However, a basic idea of TD learning is that future rewards are included in the value function. Let w^{m-1} be an estimate of w at iteration m, before a least squares update is made. Then, we can use $\bar{V}(S,x;w^{m-1})$ to obtain a value estimate of the action taken at the successive state, such that

$$w^m = \left(\sum_{n=1}^{N-1} \Phi(S^n,x^n)\Phi^\top(S^n,x^n) \right)^{-1} \times \left(\sum_{n=1}^{N-1} \Phi(S^n,x^n)\big(r(S^n,x^n,S^{n+1}) + \gamma \bar{V}(S^{n+1},x^{n+1};w^{m-1})\big) \right). \tag{25}$$

By repeating the least squares update over M iterations while collecting new samples, we can approximate the value function associated with the given policy. This algorithm is known as *least squares policy evaluation* (Nedic and Bertsekas, 2003). In the next section, we are going to present an approximate policy iteration algorithm which uses this method for policy improvement.

3.4 Approximate Policy Iteration

The idea of combining least squares updates with policy iteration has been first proposed in Lagoudakis and Parr (2003). The authors use the approximate value function of a given policy to compute an improved policy which is then used to construct a new approximate value function. The least squares approximate policy iteration (LSAPI) algorithm used in this work is shown in Figure 1.

(1) Input arguments: approximate value function $\bar{V}(\,\cdot\,; w^0)$; initial state S

(2) Define function $z(m,k) = \big((m-1)N + k \bmod D\big) + 1$

(3) Do for $m = 1, 2, \ldots, M$

  (3.1) Do for $n = 0, 1, \ldots, N-1$

    (3.1.1) $x \leftarrow \pi^E(S)$

    (3.1.2) $L_{z(m,n)} \leftarrow (S, x)$

    (3.1.3) $S' \leftarrow S^M(S, x)$

  (3.2) $w^m \leftarrow \Big(\sum_{d=z(m,1)}^{z(m,D-1)} \Phi(S_d, x_d)\Phi^\top(S_d, x_d)\Big)^{-1}$
  $\times\Big(\sum_{d=z(m,1)}^{z(m,D-1)} \Phi(S_d, x_d)\big(r(S_d, x_d, S_{d+1}) + \gamma\bar{V}(S^{d+1}, x^{d+1}; w^{m-1})\big)\Big)$

(4) Return approximate value function $\bar{V}(\,\cdot\,; w^M)$

Figure 1: Least squares approximate policy iteration

The algorithm performs M value function updates, i.e., policy improvement steps. At each iteration m, it generates a sample of N state transitions by following policy π^E and then updates the weight vector. A simple random exploration policy is sufficient, e.g., epsilon-greedy exploration, since the action space is only one-dimensional. The function $S^M(S, x)$ implements the simulation model of the Markov decision process which produces a new state S' for a given state and action. The set $L = \{(S, x, r)_1, \ldots, (S, x, r)_D\}$ represents a circular list implementation which stores a set of D state-action tuples that have been sampled sequentially. The least squares update is then made over the entire set. However, to speed up learning and ensure stability of the least squares update, only a fraction of the entire set is being changed at each iteration m, i.e., $D = kN$ with $1 \leq k \leq M$, $k \in \mathbb{N}$.

4 Numerical Results

Our research goals are to study the performance of the approximation algorithm for different parameter configurations and to analyze the influence of changes in model parameters on discounted rewards.

4.1 Experimental Design

As experimental design, we adopted a so-called *space filling design* which samples not only at the edges of the hypercube which spans the experimental area but also at its interior. We generated

#	Parameter	Default Value	Lower Bound	Upper Bound
1	Supply Std Deviation (σ_Y)	2	1	4
2	Supply Autocorrelation (θ_Y)	0.5	0.0	0.75
3	Price Std Deviation (σ_P)	2	1	4
4	Price Autocorrelation (θ_P)	0.5	0.0	0.75
5	Correlation of Y and P (ρ_{PY})	-0.5	-0.75	0.0
6	Storage Capacity (C)	4	0	10
7	Storage Efficiency (η)	1	0.5	1
8	Discount Rate (γ)	0.9	0.0	0.9
9	Negative Imbalance Price Factor (u)	1.0	0.5	2.0
10	Positive Imbalance Price Factor (o)	0.0	0.0	0.5

Table 1: Default parameters and parameter ranges for the experimental design

a low-discrepancy *Faure* sequence to select design points which are uniformly scattered over the design space. Faure sequences have the useful property that a longer sequence can be constructed from a shorter one by resuming the sequence while still preserving uniformity, see Chen et al. (2006) for a review on designs for computer experiments.

As the basic setup for our studies we used default values and parameter intervals as shown in Table 3.1. Mean supply and mean price are fixed to $\mu_Y = \mu_P = 5$, because their magnitude only influences the volume but not the structure of the bids as long as the ratios of mean, variance and capacity are held constant. The storage efficiency is defined by its round-trip efficiency, $\eta = \eta^+\eta^-$.

To ensure tractability of the linear program, the stochastic processes of Y and P were bounded to the discrete set $\{0, 1, ..., 10\}$. Accordingly, the continuous counterparts were truncated at zero and 2μ to ensure comparability.

The parameters of LSAPI were pre-optimized so that the algorithm converges. The algorithm was set to collect batches of $T = 500$ samples and performs $M = 200$ value function updates. All samples were stored in a circular list of size $N = 25'000$. As exploration policy the algorithm used a simple epsilon-greedy policy, where a random bid is chosen with probability $\epsilon = 0.01$ and policy π is executed with probability $1 - \epsilon$.

We implemented the proposed algorithms in Java with Matrix computations being done by the Jama package. The linear programming approach was implemented and solved with FICO Xpress 7 and linked to Java via the XPRM model interface. All statistical analyses were done with SPSS 17.

Policy	Algorithm	Mean Reward	Mean Difference	
Optimal Discrete Policy	LP	19.94	LP	LSAPI-2
Second-order Polynomial	LSAPI-2	20.63	0.69 ± 0.31	–
Third-order Polynomial	LSAPI-3	20.74	0.80 ± 0.33	(0.11 ± 0.14)

Table 2: Mean rewards and 95% confidence intervals of the difference

4.2 Solution Quality

To test the quality of the approximation method, we benchmarked the policies of two value function approximations against policies computed with linear programming (LP). To use the optimal discrete policy during simulation, we projected the continuous states into the discrete state space by rounding to the next integer. Also, the round-trip efficiency was set to $\eta = 1.0$ to ensure comparability. For $\eta < 1.0$, we can expect the discrete policy to perform even worse, due to the additional rounding error.

The Weierstrass approximation theorem states that every continuous function defined on a closed interval can be approximated by a polynomial function to any degree of accuracy. Denote $z = \{x, y, p, g\}$ and $\bar{V}_2(S, x)$ and $\bar{V}_3(S, x)$ as second-order and third-order polynomial value function approximations, such that

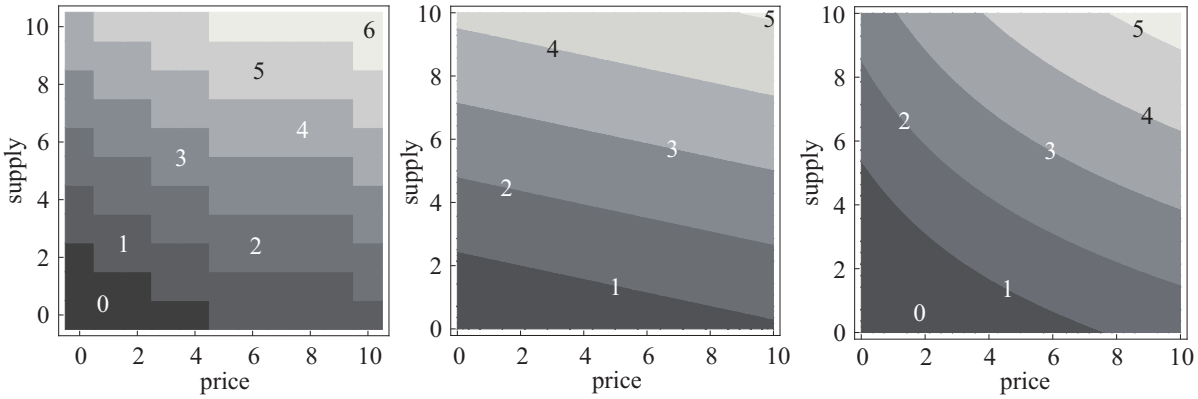
$$\bar{V}_2(S, x) = w_0 + \sum_{i=1}^{4} w_i z_i + \sum_{i=1}^{4} \sum_{j=i}^{4} w_{ij} z_i z_j, \tag{26}$$

$$\bar{V}_3(S, x) = w_0 + \sum_{i=1}^{4} w_i z_i + \sum_{i=1}^{4} \sum_{j=i}^{4} w_{ij} z_i z_j + \sum_{i=1}^{4} \sum_{j=i}^{4} \sum_{k=j}^{4} w_{ijk} z_i z_j z_k. \tag{27}$$

Let LSAPI-2 and LSAPI-3 be the corresponding ADP algorithms. For a large enough sample size, we expect LSAPI-3 to compute a more accurate approximation than LSAPI-2 because $\bar{V}_2(S, x)$ is contained in $\bar{V}_2(S, x)$.

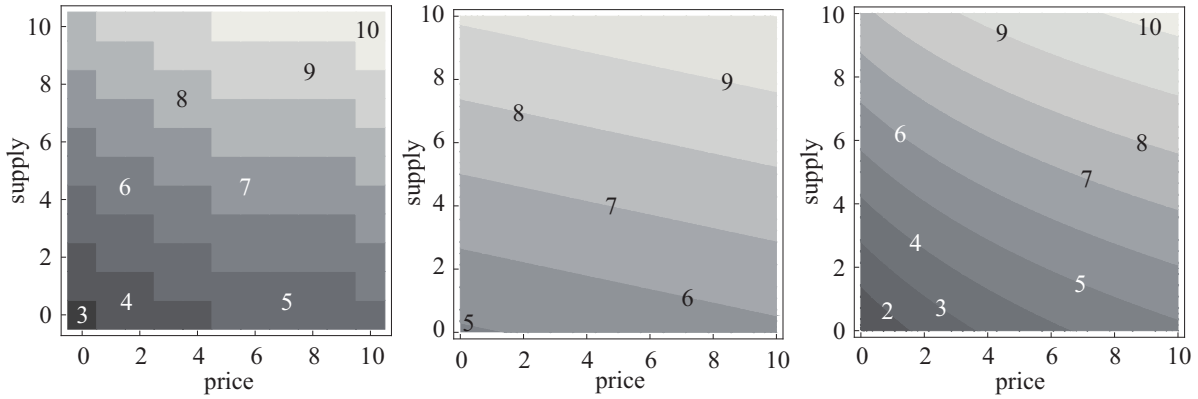
We ran the Faure sequence to generate 90 model configurations uniformly distributed over the parameter space. LSAPI-2, LSAPI-3 and LP then approximated an optimal policy for each configuration. The average reward from following each policy is recorded during simulation. The results are summarized in Table 2.

The table shows that LSAPI was able to approximate the optimal bidding strategy and delivers sufficient policies with both value function approximations. The LSAPI-2 policy achieved a 3.5% ($p < .01$) higher reward than the LP policy, and the LSAPI-3 policy achieved a 4.0% ($p < .01$) higher reward than the LP policy. The difference between LSAPI-2 and LSAPI-3,



(a) LP (b) LSAPI-2 (c) LSAPI-3

Figure 2: Contour plots of the approximate policies at $g = 0$



(a) LP (b) LSAPI-2 (c) LSAPI-3

Figure 3: Contour plots of the approximate policies at $g = 4$

however, is not significant.

4.3 Policy Analysis

For illustrative purposes, we plot the different policies produced by LSAPI-2, LSAPI-3 and the LP for the default parameter configuration. Since the state-action space is four-dimensional, we fixed the storage state and drew contour plots of the surface which maps inputs of price and supply to bidding decisions, as suggested by the respective policies. Figures 2 and 3 show the contour plots of the three value functions for storage state $g = 0$ and $g = 4$, respectively.

The contour plot of the LP policy is non-smooth and areas with equal elevation, i.e., bidding decision, are shaded in the same color. For the two LSAPI policies, contour lines are drawn along integer elevations. Although the contour plots do not look alike, they share important similarities. Across all three policies, the elevation around the means of price and supply is nearly identical. The differences among the policies become larger towards the edges of the

graph where state transition probabilities are lower and the associated errors have less impact on policy performance. The slopes of the contour lines exhibit a comparable inclination, so that bids respond to an increase in price and supply.

Note that the contour lines of the LP and LSAPI-3 policies share a comparable curvature. The LSAPI-2 policy, on the other hand, does not capture this detail because the derivative of its value function is linear in price, supply and storage. This leads to an increasing difference between LSAPI-2 and LSAPI-3/LP towards the edges of the graph where the state transition probabilities become lower. LSAPI-2 puts more weight on states which are frequently visited and balances the errors which emerge from decisions in extreme states.

We conclude that using simple polynomials to approximate the value function of the continuous state MDP yields excellent results. The corresponding policies performed even better than the optimal policy of the discrete counterpart.

4.4 Metamodel Analysis

To study the sensitivity of the objective value towards changes in model parameters, we analyze the model using regression analysis. We ran the model with 2000 different parameter configurations generated by the space filling design. To approximate the corresponding optimal rewards, we used LSAPI-2 as it produces high-quality policies at low computational cost.

Denote Z as the set of model parameters and Z_i as the value of the i-th parameter, and let us analyze the relationship of model parameters and discounted reward with the following two regression equations:

$$\hat{Q}_1 = \beta_0 + \beta_1 Z_1 + \beta_3 Z_3 + \beta_5 Z_5 + \beta_6 Z_6 + \beta_9 Z_9 + \beta_{10} Z_{10} \tag{28}$$

$$\hat{Q}_2 = \hat{Q}_1 + \beta_{12} Z_{12} + \beta_{345} Z_{345} + \sum_{i=1}^{10} \beta_{i6} Z_i Z_6 \tag{29}$$

The first equation captures the main effects of model parameters which have a direct influence on the response variable, i.e., the impact of a parameter change on the discounted reward obtained by following the LSAPI-2 policy. The second equation additionally includes interactions of model parameters with storage capacity Z_6 as well as two interaction terms to account for the indirect effect of autocorrelation. For both models, a summary of the regressions analysis is given in Table 3.

Table 4 shows the results from running a regression on $\tilde{Q}_1$. The explanatory power of this

Model	R^2	R^2_{adj}	Std Error
$\hat{Q}_1$	0.530	0.528	1.900
$\hat{Q}_2$	0.898	0.897	0.889

Table 3: Model summary

$\hat{Q}_1$	Main Effect	Coefficient	Std Error	t-Statistic	p-Value
Z_0	(Constant)	24.719	0.252	97.917	0.000
Z_1	Supply Std Deviation	-2.008	0.049	-40.924	0.000
Z_3	Price Std Deviation	-0.296	0.049	-6.040	0.000
Z_5	Correlation of Y and P	2.661	0.294	9.038	0.000
Z_6	Storage Capacity	0.274	0.015	18.614	0.000
Z_9	Negative Imbalance Price Factor	-0.805	0.098	-8.206	0.000
Z_{10}	Positive Imbalance Price Factor	1.636	0.294	5.555	0.000

Table 4: Coefficients of the main effects

model is poor ($R^2 = 0.53$) because it only accounts for the individual effects of variability, storage capacity, correlation and imbalance costs on discounted rewards.

We find that supply and price deviation both have a negative effect on rewards ($\beta_1 < 0$, $\beta_3 < 0$), because higher variability increases the costs of imbalances. The difference in magnitude between both effects is substantial. This asymmetry originates in the nature of the stochastic processes with the unilateral dependency of price on supply. Changes in the standard deviation of the supply process therefore affect the price process but not vice versa, so that the overall impact of price variability is lower than the impact of supply variability. However, the correlation of price and supply increases rewards ($\beta_5 > 0$), which may be an argument in favor of investments in solar power which has the maximum yield when consumption is up, i.e., during daytime and summertime in warmer climates. As a matter of course, a higher storage capacity and price factor for positive imbalance have a positive effect on rewards ($\beta_6 > 0$, $\beta_{10} > 0$) while a higher price factor for negative imbalance has a negative effect ($\beta_9 < 0$).

Table 3.5 reports the results from running a regression on $\tilde{Q}_2$, but shows only the interaction terms. A high sample correlation ($R^2 = 0.898$) indicates that the second model already delivers a relatively accurate prediction of the relationship between model parameters and rewards.

Theoretically, supply and price autocorrelation do not have a direct effect on discounted rewards, because they explain the variability of the stochastic process only to some extent. Therefore, we study the interaction of supply autocorrelation with supply standard deviation. In contrast to the supply process, prices moreover depend on their correlation with supplies, so that we need to study the three-way interaction of price variability, autocorrelation and supply-

$\hat{Q}_2 \setminus \hat{Q}_1$	Interaction Effect	Coefficient	Std Error	t-Statistic	p-Value
$Z_1 Z_2$	Supply Autocor × Supply Std Dev	1.525	0.057	26.986	0.000
$Z_3 Z_4 Z_5$	Price Autocor × Price Std Dev × Cor	0.899	0.148	6.080	0.000
$Z_1 Z_6$	Capacity × Supply Std Deviation	0.156	0.008	19.561	0.000
$Z_2 Z_6$	Capacity × Supply Autocorrelation	-0.162	0.026	-6.236	0.000
$Z_3 Z_6$	Capacity × Price Std Deviation	0.019	0.008	2.399	0.017
$Z_4 Z_6$	Capacity × Price Autocorrelation	-0.134	0.020	-6.577	0.000
$Z_5 Z_6$	Capacity × Correlation of Y and P	-0.285	0.048	-5.967	0.000
$Z_6 Z_6$	Capacity × Capacity	-0.063	0.003	-23.432	0.000
$Z_7 Z_6$	Capacity × Storage Efficiency	1.391	0.024	58.343	0.000
$Z_8 Z_6$	Capacity × Discount Rate	0.348	0.013	26.265	0.000
$Z_9 Z_6$	Capacity × Negative Imbalance Price	0.233	0.016	14.692	0.000
$Z_{10} Z_6$	Capacity × Positive Imbalance Price	-0.581	0.048	-12.163	0.000

Table 5: Coefficients of the interaction effects

price correlation. In both cases, we find that autocorrelation decreases the negative impact of variability ($\beta_{12} > 0$, $\beta_{345} > 0$), as it stabilizes the time series and thereby decreases the risk of imbalances.

Two-way interactions of model parameters with storage capacity uncover the individual impact of parameter changes on the contribution of storage capacity to rewards. While the value of storage increases in standard deviation ($\beta_{16} > 0$, $\beta_{36} > 0$), supply and price autocorrelation decrease the value of storage ($\beta_{26} < 0$, $\beta_{46} < 0$). This effect can be explained by comparing two stochastic processes with identical first two moments, one with positive autocorrelation and one without. The process with positive autocorrelation will exhibit a wider average amplitude than the process without. With wider amplitudes we need more storage to buffer the same variability, which decreases the per unit value of storage capacity. The effect of autocorrelation is strong in a renewable power portfolio with only one power source. To decrease its unfavorable effect, investors should diversify, e.g., combine wind and solar power or invest cross-regionally.

Furthermore, we expect the marginal value of storage to decrease due to diseconomies of scale. The negative quadratic effect ($\beta_{66} < 0$) of storage capacity provides evidence for a concave function of storage value on storage capacity. Storage efficiency, on the other hand, increases the value of storage ($\beta_{76} > 0$).

From a technical point of view, the discount factor enables the approximation method to develop a bidding strategy which anticipates future states and decisions. Without a discount factor ($\gamma = 0.0$), there is no need for policy iteration and a one-shot least squares estimation as in (24) would be sufficient. This would produce a myopic policy which maximizes the second-stage reward. Figure 4 shows the average reward, i.e., equivalent annuity value, for the default model

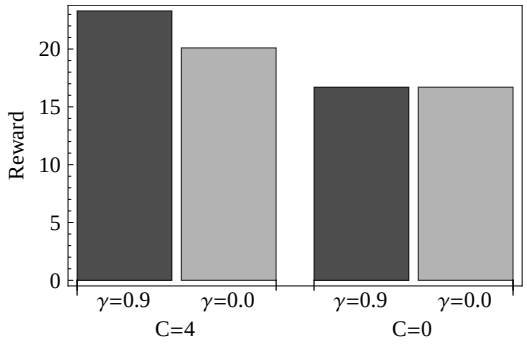


Figure 4: Rewards from optimal and myopic policies for systems with and without storage

configuration from using a multistage policy ($\gamma = 0.9$) versus using a myopic policy ($\gamma = 0.0$). For the system with storage ($C = 4$), the multistage policy yields a higher average reward than the myopic policy. A power producer with a hybrid system would therefore benefit from using a policy which anticipates future states and decisions. This also explains why the value of capacity increases in the discount factor ($\beta_{86} > 0$). For the system without storage ($C = 0$), however, multistage and myopic policy yield identical rewards, because there is no storage balance which links successive periods. In that case, a two-stage approach similar to the one proposed in Matevosyan and Söder (2006) is optimal.

Figure 4 also shows that for the given model configuration rewards of the system without storage are significantly lower than rewards of the hybrid system. We conclude that the financial benefit of storage is twofold: first, there is the risk associated with imbalance costs. A lower price for positive imbalance as well as a higher price for negative imbalance both increases the value of storage ($\beta_{96} > 0$, $\beta_{10,6} < 0$). Second, storage has the ability to take advantage of price arbitrage and alleviates the necessity to sell power when the price is low. Accordingly, the value of storage increases in negative correlation of price and supply ($\beta_{56} < 0$).

5 Conclusion and Outlook

In this paper, we presented a model of the optimal bidding strategy for renewable power generation with storage option at a day-ahead market. We formulated the model as a continuous-state Markov decision process and presented a solution approach based on approximate dynamic programming. We used least squares policy evaluation within the approximate policy iteration framework to approximate the value function of the optimal policy. For policy evaluation and improvement, the algorithm had access to a simulation model of the process. As a benchmark, we computed the transition matrix of the stochastic decision process for a small discrete

approximation of the state space and used linear programming to determine an optimal policy.

We found that approximate value functions based on simple polynomials yield better policies for the continuous state space than the optimal policy of a discretization. This effect will become even more profound when rounding errors occur due to storage efficiency losses.

A numerical analysis of the response surface of rewards on model parameters revealed that supply uncertainty, imbalance costs and a negative correlation of market price and supplies are the main drivers for investments in storage. An interesting result is that the value of storage decreases in autocorrelation, as more capacity is needed to buffer a stochastic process with a wider amplitude.

References

Bathurst, G. N. and G. Strbac (2003). Value of combining energy storage and wind in short-term energy and balancing markets. *Electric Power Systems Research 67*, 1–8.

Bathurst, G. N., J. Weatherill, and G. Strbac (2002). Trading wind generation in short term energy markets. *IEEE Transactions on Power Systems 17*(3), 782–789.

Brunetto, C. and G. Tina (2007). Optimal hydrogen storage sizing for wind power plants in day ahead electricity markets. *IET Renewable Power Generation 1*(4), 220–226.

Chen, C. C. P., K.-L. Tsui, R. R. Barton, and M. Meckesheimer (2006). A review on design, modeling and applications of computer experiments. *IIE Transactions 38*(4), 273–291.

Fleten, S. E. and T. K. Kristoffersen (2007). Stochastic programming for optimizing bidding strategies of a nordic hydropower producer. *European Journal of Operational Research 181*, 916–928.

García-González, J., R. M. R. De la Muela, L. M. Santos, and A. M. González (2008). Stochastic joint optimization of wind generation and pumped-storage units in an electricity market. *IEEE Transactions on Power Systems 23*(2), 460–468.

Graves, F., T. Jenkin, and D. Murphy (1999). Opportunities for electricity storage in deregulating markets. *The Electricity Journal 12*(8), 46–56.

Korpås, M. and A. T. Holen (2006). Operation planning of hydrogen storage connected to

wind power operating in a power market. *IEEE Transactions on Energy Conversion 21*(3), 599–606.

Lagoudakis, M. G. and R. Parr (2003). Least-squares policy iteration. *Journal of Machine Learning Research 4*, 1107–1149.

Matevosyan, J. and L. Söder (2006). Minimization of imbalance cost trading wind power on the short-term power market. *IEEE Transactions on Power Systems 21*(3), 1396–1404.

Nedic, A. and D. P. Bertsekas (2003). Least squares policy evaluation algorithms with linear function approximation. *Discrete Event Dynamic Systems 13*, 79–110.

Neubarth, J., O. Woll, C. Weber, and M. Gerecht (2006). Beeinflussung der Spotmarktpreise durch Windstromerzeugung. *Energiewirtschaftliche Tagesfragen 56*(7), 42–45.

Puterman, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.

Schainker, R. B. (2004). Executive overview: energy storage options for a sustainable energy future. In *IEEE Power Engineering Society General Meeting*, pp. 2309–2314. IEEE.

Sutton, R. S. and A. G. Barto (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Wen, F. and A. K. David (2000). Strategic bidding in competitive electricity markets: a literature survey. In *IEEE Power Engineering Society Summer Meeting*, pp. 2168–2173.