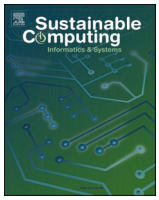
Sustainable Computing: Informatics and Systems 29 (2021) 100477

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

journal homepage: www.elsevier.com/locate/suscom



A Probabilistic Approach to Committing Solar Energy in Day-ahead Electricity Markets

Noman Bashir *, David Irwin, Prashant Shenoy

University of Massachusetts Amherst, USA

ARTICLE INFO

Keywords:
Solar energy
Forecasting
Cloud cover
Day-ahead electricity markets

ABSTRACT

Grid-tied solar is governed by a variety of complex regulations. Since a higher solar penetration imposes indirect costs on the grid, these regulations generally limit the aggregate amount of grid-tied solar, as well as the compensation its owners receive. These regulations are also increasingly limiting solar's natural growth by preventing users from connecting it to the grid. One way to address the problem is to partially deregulate solar by allowing some solar generators to participate in the electricity market. However, day-ahead electricity markets require participants to commit to selling energy one day in advance to ensure system stability and avoid price volatility. Thus, to operate in the day-ahead market, solar generators must solve a *solar commitment problem* by determining how much solar energy to commit to sell each hour of the next day that maximizes their revenue despite the uncertainty in next-day solar generation. We present a probabilistic approach to addressing the solar commitment problem that combines a solar performance model with an analysis of weather measurement and forecast data to determine a conditional probability distribution over next-day solar generation outcomes, which we use to determine solar energy commitments each hour that maximize expected revenue. We show that, as the deviation penalty for over-committing solar increases, our probabilistic approach enables increasingly more savings than a deterministic approach that simply trusts weather measurements and forecasts.

1. Introduction

The cost of solar power is steadily declining due to both continuing improvements in solar cell efficiency and economies of scale in manufacturing. These cost decreases have made solar competitive with traditional fossil fuel-based generation in many areas. In addition to cost, solar offers numerous other benefits compared to traditional generation sources, including zero carbon emissions, low operating costs, few maintenance requirements, and incremental scalability, which enables solar operators to scale up capacity by adding more modules. As a result, solar penetration in the grid, i.e., the fraction of grid energy that derives from solar, is rapidly increasing. However, this increase is being closely regulated by governments, utility boards, and independent system operators (ISOs), as large-scale solar adoption is altering the grid's economic and operational model.

With respect to the grid's economic model, increased solar adoption is democratizing energy generation by enabling smaller-scale distributed generation to compete with large-scale centralized generation. Unlike traditional fuel-based generators, which become significantly more energy- and cost-efficient at larger scales, solar costs scale more linearly, such that the cost per capacity of even small solar farms or residential deployments is competitive with larger-scale solar farms. High solar penetrations also alter the grid's operational model, which assumes i) a highly predictable demand that primarily varies with temperature and ii) a controllable supply. Grid operators typically determine generator dispatch schedules one day in advance based on predictable load forecasts to ensure there is enough supply online to satisfy the next day's expected demand. Since solar is treated as an uncontrollable supply that, as we discuss, is variable and challenging to forecast, it complicates such day-ahead generator dispatch scheduling which reduces the grid's reliability.

Due to the issues above, grid-tied solar is highly regulated. In the United States, these regulations vary widely by state and are based, in part, on policy objectives. For example, some states, such as California and Massachusetts, have generous policies that encourage, and partially subsidize, grid-tied solar to increase adoption, while other states have policies that discourage, and effectively penalize, grid-tied solar. In many cases, solar connection policies place limits on the solar capacity

* Corresponding author.
E-mail addresses: nbashir@umass.edu (N. Bashir), deirwin@umass.edu (D. Irwin), shenoy@cs.umass.edu (P. Shenoy).

https://doi.org/10.1016/j.suscom.2020.100477
Received 9 June 2020; Received in revised form 4 October 2020; Accepted 18 October 2020
Available online 2 November 2020

that can connect to the grid, and the compensation owners receive for the energy it generates. Importantly, these policies generally *do not* require solar generators to compete with traditional sources in the electricity market. Instead, operators feed all solar energy into the grid, and receive some pre-determined compensation for it, e.g., the retail rate. Unfortunately, this approach is impeding solar's natural growth—by preventing new solar connections once solar capacity hits its limit—and does not properly incentivize operators to address the unpredictability of solar.

An alternative approach is to relax restrictions on connecting solar to the grid—by allowing anyone to connect—but then partially deregulate solar by allowing some solar generators to participate in the same electricity markets as other generators. Some regions are already experimenting with such deregulation and competition [1]. ISOs typically operate a day-ahead hourly market that requires participants to *commit* to selling wholesale electricity one day in advance to ensure system stability and avoid price volatility. Market participants that do not meet their commitments must then pay a *deviation penalty* based on the difference between the energy they supplied and their commitment [2]. From the grid's perspective, similar solar commitments would simplify operations by providing accurate knowledge of solar generation capacities for the following day, effectively turning uncontrollable generation sources into ones that appear to be controllable. From the solar operator's perspective, however, due to the uncertainties in next-day generation, optimally committing solar energy in the day-ahead market is challenging, as it requires operators to determine their commitment level based on uncertain solar forecasts and the potential penalty. Of course, more accurate solar forecasts directly translate to more accurate energy commitments, fewer penalties, and higher revenue. Thus, this approach directly incentivizes more accurate solar forecasting, and incorporates solar's unpredictability into the revenue it generates.

Ultimately, the accuracy of day-ahead solar forecasting is based on weather forecast accuracy, specifically for temperature and cloud cover, which are the primary weather metrics that affect solar generation [3]. In particular, the National Weather Service (NWS) in the United States runs numerical weather prediction (NWP) models for every 2.5km^2 area of the country in real-time, and releases them to the public. The National Digital Forecast Database (NDFD) includes archives of these forecasts going back over a decade.[1] Unfortunately, unlike temperature measurements and forecasts, cloud cover measurements and forecasts are often much more imprecise, which limits day-ahead solar forecasting accuracy, even assuming a solar performance model that is otherwise perfect.

To address the problem, we present a probabilistic approach that considers cloud cover measurement and forecast uncertainty in determining day-ahead solar commitments to optimize revenue in day-ahead electricity markets. Our approach combines a solar performance model with an analysis of public cloud cover measurement and forecast data to determine a *probability distribution* over possible day-ahead generation outcomes, which we then use to determine day-ahead solar energy commitments each hour that maximize expected revenue. Our approach is complementary to using energy storage to reduce the variations in, and improve the predictability of, renewable energy fed into the grid, as done in prior work in the context of wind energy generation and electricity markets [4]. We evaluate the potential of energy storage to increase the revenue of solar generators that compete in the market in §5.5.

Our hypothesis is that a probabilistic approach to determining solar commitments that considers cloud cover measurement and forecast uncertainty will yield more revenue than a deterministic approach that simply trusts the measurements and forecasts. In evaluating our hypothesis, we make the following contributions.

Cloud Cover Data Analysis. We analyze cloud cover measurements and forecasts from publicly-available weather data released by the NWS and NDFD to quantify their uncertainty. Our analysis results in probability distributions of solar energy output for cloud cover measurements and forecasts under different conditions.

Probabilistic Solar Commitments. We present a probabilistic approach to committing solar in the day-ahead market. Our approach leverages the probability distributions above to compute the expected revenue for different day-ahead solar commitments based on the day-ahead electricity price, the penalty for over-commitments, and the loss of revenue from under-commitments. We then select the solar commitment that maximizes expected revenue, given the uncertainty.

Implementation and Evaluation. We implement our probabilistic approach by extending a publicly-available open-source solar performance model to include uncertainty in cloud cover measurements and forecasts [3]. We evaluate our approach using price data from a regional ISO's day-ahead energy market, and show that, as the penalty for over-committing solar increases, our probabilistic approach enables increasingly more savings than a deterministic approach that simply trusts cloud cover measurements and forecasts.

2. Background

Below, we provide background on electricity markets, cloud cover measurements and forecasts, and solar forecasting.

2.1. Electricity Markets

As discussed in §1, the day-ahead electricity market requires participants to commit to provide a certain amount of energy each hour of the next day at a certain price. To determine the commitments and price, participants submit bids at a certain time, e.g., between 8am and 10am, one day in advance that specify an energy commitment and price for each hour of the next day. The market operator sets the price each hour using a uniform price multi-unit (or "dutch") auction, such that all bidders are paid the price of the highest winning non-zero bid, where the number of units of energy sold each hour is based on the next day's load forecast for that hour [5]. The per-hour next-day load forecasts are highly accurate (in the absence of behind-the-meter grid-tied solar), since they are largely based on accurate temperature forecasts and well-known societal patterns, e.g., holidays, weekends, and weekdays. Traditional fuel-based generators generally bid a price based on their fuel cost, as generating more energy burns more fuel, which increases cost. In contrast, since solar operators have no fuel cost and few other operating costs, their optimal bid is at or near $0. Thus, solar operators are *price takers* in that they are willing to sell energy for effectively any positive price [6]. Note that nuclear power plants and hydroelectric plants are also price takers in the current market, since increasing their energy production also does not increase their operating costs.

Thus, when operating in the day-ahead market, solar operators only need to determine the energy to commit each hour of the next day, and not a bid price. The amount of energy that maximizes expected revenue is a function of the solar forecast accuracy, the day-ahead price, and the deviation penalty incurred for over-committing solar energy due to an inaccurate forecast, each of which introduce uncertainty. Note that the deviation penalty operators pay is in addition to the cost they incur for buying energy in the 5-minute spot market to make up for their energy deficit. However, when day-ahead load forecasts are accurate, the spot price and day-ahead converge to the same value. As a result, on average, operators do not incur a loss (or gain) from being forced to buy energy in the spot market to make up for over-commitments [2]. Thus, we only consider the deviation penalty when assessing the negative impact of over-committing solar energy in the market. In practice, deviation penalties range from as low as 2% to over 100% of the day-ahead price [2]. We evaluate our work for a wide range of deviation penalties in §5.

[1] The NDFD's forecasts are slightly different, as they include adjustments by humans.

2.2. Weather Measurements and Forecasts

Weather forecasts derive from NWP models, such as the Weather Research and Forecast (WRF) model [7], run on supercomputers by government agencies, such as the NWS in the United States, to benefit the public. However, imperfect models and uncertainties in initial atmospheric conditions limit forecast accuracy. Weather forecasts include predictions for numerous weather metrics. Temperature and cloud cover are the primary weather metrics that affect solar generation and forecasting.[2] In general, temperature changes slowly based on the movement of weather fronts and the day/night cycle, which results in accurate day-ahead predictions generally within ±5C. Since solar conversion efficiency decreases by only ∼0.5% for each 1C change in temperature, the small inaccuracy in temperature forecasts does not significantly impact solar forecast accuracy.

In contrast, cloud cover can be highly variable, changing frequently over a given location as clouds pass by. As a result, cloud cover measurements and forecasts are much less reliable. While NWP models generally forecast cloud cover as a percentage, these percentages are often mapped to five levels of cloudiness in widely-available public weather data and forecasts, such as from common weather sites like Weather Underground. These five levels include clear (CLR), few (FEW), scattered (SCT), broken (BKN), and overcast (OVC). Each level corresponds to ground-level cloud cover measurements quantified in *oktas*, where an okta represents one-eighth of the sky [8]. In NWS forecasts, CLR translates to <1 okta, FEW to 1-3 oktas, SCT to 3-5 oktas, BKN to 5-7 oktas, and OVC to 8 oktas. Some sources, such as ASOS [9], define slightly different okta translations. In this paper, we use METAR translations released by the NWS [8,10]. Note that neither the raw cloud cover percentage generated by NWP models nor okta measurements account for the effect of specific cloud characteristics, such as cloud type, height, or thickness, and thus both are coarse and imprecise measures of cloud cover. In addition, cloud cover is reported for small regions based on observations at a single location, yet clouds are often small and move fast, such that cloud cover can differ at nearby locations.

In general, cloud cover is more difficult to measure compared to other weather metrics, particularly rain and snow, in part because remote sensing of clouds is difficult. Unlike rain and snow, weather radars cannot sense clouds. While cloud cover estimates are possible from visible satellite imagery, these images only capture the tops of clouds and, like ground-level measurements, do not account for cloud depth, height, or thickness. As a result, cloud cover estimates from visible satellite imagery are not necessarily better than ground-level observations, as at least the latter captures what is happening on the ground. Since existing cloud cover measurements and forecasts are coarse and imprecise, they introduce inherent uncertainty into forecasting any site's solar energy output.

Note that the NWS processes the output of sophisticated numerical weather prediction models to generate a cloud cover forecast. The National Digital Forecast Database (NDFD) makes historical archives of these weather forecasts released every hour (with each covering the next 168 hours) for every location in the United States for multiple decades [11]. These forecasts are also released in real time every hour by the NWS. We focus on extracting and leveraging cloud cover measurement and forecast uncertainty from this public archival data in determining solar energy commitments.

2.3. Solar Performance Modeling and Forecasting

There is substantial work on solar energy forecasting using different input data, at different time resolutions, e.g., hourly, daily, or annually, and for different time horizons, e.g., hour-ahead or day-ahead [12]. The simplest forecast model is a persistence model that predicts all future time periods, of any resolution, will have the same value as the present time period. Persistence models for solar energy forecasting perform well over short time periods, e.g., less than 2 hours, or long resolutions, e.g., annually. There has also been substantial work on analyzing satellite images to estimate near-term cloud movement to forecast solar over short time horizons, e.g., 3-4 hours [13]. However, this approach does not work for day-ahead forecasting. Instead, day-ahead solar energy forecasting must use the weather forecasts above as input to a solar performance model that estimates solar output based on the forecast, as well as a site's location, physical characteristics, time-of-day, and day-of-year. Note there has also been work on using NWP models to deterministically forecast solar radiation [14]. We extend this work by using historical archives of these NWP forecast models to determine the location-specific uncertainty in these forecasts based on the current conditions when applied to solar energy (rather than solar radiation).

Since we focus on solar energy forecasting, we must infer solar output based on current conditions. There are many solar performance models available that do this. We leverage a "black box" model, called Solar-TK, that is open-source and publicly available. This model calibrates its parameters entirely from a small amount of historical generation data from a solar site [3]. Once calibrated, the model only requires as input a site's location, time-of-day, day-of-year, cloud cover, and temperature over some time resolution, and returns as output an estimate of solar energy over that time resolution. Any model's primary source of inaccuracy and uncertainty is the coarse cloud cover measurements and forecasts described in §2.2. That is, all existing models are highly accurate at inferring a solar site's generation under clear skies and a known temperature, time-of-day, and day-of-year. An advantage of using an open-source model based on publicly available data is that our approach can be directly applied to any solar site in the U.S. Note that Solar-TK's model already generates a deterministic solar forecast. Our work extends this model to generate a probabilistic forecast, which we apply to the solar commitment problem. We compare our probabilistic forecasting approach with a deterministic one in §5.

3. A Probabilistic Approach

Our work addresses the problem of determining how much solar energy to commit in the day-ahead electricity market each hour of the next day to maximize revenue. Over-committing solar energy incurs a financial penalty and under-committing it requires potentially wasting energy by actively curtailing solar output, e.g., using smart inverters and power optimizers [15], and losing the associated revenue. To address the problem, we take an empirical approach that incorporates uncertainty in cloud cover measurements and forecasts when determining the solar energy to commit to optimize revenue.

We use historical cloud cover measurement and forecast data to determine an error distribution for each. The inaccuracy in cloud cover measurements is universal, and thus we compute it across many solar sites and time periods in terms of a normalized *clear sky photovoltaic (PV) index*, or K_{pv}, which represents the fraction of power a solar site generates relative to the maximum power it could have generated at the same time but under clear skies [16,17]. Computing K_{pv} requires knowing a site's maximum solar generation potential under clear skies at any point in time. Solar-TK [3] discussed in §2.3 accurately models this maximum solar generation.

K_{pv} normalizes for differences between solar sites and time periods. As a result, K_{pv} is same for all solar sites operating under the exact same cloud cover conditions, regardless of their location, physical characteristics, time-of-day, or day-of-year. Thus, if cloud cover measurements were perfectly accurate then the same measurement value should result in the same K_{pv}. However, as we show, K_{pv} varies significantly for the same cloud cover measurement. These measurement errors are expected based on the discussion in §2.

Note that, while K_{pv} is a continuous metric that can take on any value

[2] We do not consider snow, which also affects solar in some locations.

between 0 and 1, we only consider the 5 high-level okta-based cloud cover levels from the previous section. As we show, there is significant inaccuracy even after summarizing NWP model output, which specifies the percentage of cloud cover, using these coarse levels. Ultimately, for any measured cloud cover, the actual position of the specific clouds in the sky relative to the Sun's position, as well as their characteristics, determines precisely how much solar irradiance reaches the ground, which results in different K_{pv} for the same cloud cover measurement To quantify the uncertainty in K_{pv} given a cloud cover measurement, we compute a conditional distribution from historical data that captures the probability of generating a specified range of K_{pv}'s given a particular cloud cover measurement, e.g., FEW, SCT, BKN, OVC, or CLR.

Unlike measurement accuracy, forecast accuracy is specific to each location and its local climate. That is, some locations and climates are inherently easier to forecast. As a result, we estimate a forecast error distribution using only data from a specific solar site's location. We capture the forecast error using a conditional distribution that captures the probability that a certain type of cloud cover occurs given that the forecast called for a potentially different type of cloud cover. For example, if the next day forecast calls for clear (CLR) skies at 9am, the distribution captures the probability, given a CLR forecast, of measuring each of the 5 types of cloud cover (FEW, SCT, BKN, OVC, and CLR) at 9am the next day. Again, by translating the cloud cover percentage from NWP models into these 5 summary levels, we would expect to get more accurate results.

Given distributions above, our probabilistic commitment approach accounts for measurement and forecast errors by computing the commitment amount that maximizes expected revenue. As we discuss, our approach balances a tradeoff between increased revenue from committing more solar energy, and decreased revenue from incurring penalties due to errors in measurements and forecasts.

3.1. Cloud Forecast and Measurement Uncertainty

Below, we outline how we quantify cloud cover forecast and measurement uncertainty, and its effect on estimating and forecasting solar generation, from publicly-available data.

Cloud Cover Forecast Uncertainty. NWS weather forecasts are issued every hour for every location in the U.S., and include estimates of weather metrics every 3 hours for the next week in the future. We distill next-day forecasts into hour-level forecasts, such that each hour of the 3-hour forecast has the same value. The National Digital Forecast Database (NDFD) stores archives of historical NWS weather forecasts for every location in the United States for over a decade. We quantify the accuracy of historical forecasts by comparing them with ground truth hourly weather observations, which we take from the NWS Integrated Surface Database (ISD) dataset [18] and Weather Underground [19] from 2011-2018.

As noted above, the NDFD and weather observation datasets report cloud cover as a percentage from 1-100%, which we translate to one of five strings for cloud cover from §2.2, e.g., CLR, FEW, SCT, BKN, and OVC [11]. In making this translation, we convert the cloud cover percentages in the observations to strings based on their definition, where CLR is 0-5.5% cloud cover, FEW is 5.5-25.5%, SCT is 25.5-50.5%, BKN is 50.5-87.5%, and OVC is 87.5-100%. As we show, even after this translation there is significant error in measurements and forecasts. For example, in our case, if $K_{pv} = 51\%$, as long as the NWP percentage is between 50.5-87.5%, then the K_{pv} would match the measurement or forecast. In addition, doing this translation ensures that our historical cloud cover observations and forecasts have the same hourly resolution—equivalent to the day-ahead market's resolution—and scale.

Given historical traces of a location's cloud cover observations and forecasts, we compute a conditional probability distribution that captures the forecast's error as shown below.

$$\Pr(\text{observation} = Z | \text{forecast} = Y) \quad (1)$$

The conditional probability distribution captures the probability of observing a particular type of cloud cover given a particular cloud cover forecast for that location, where Z and Y each may take one of five values corresponding to the cloud cover strings CLR, FEW, SCT, BKN and OVC. As one example, given a FEW forecast at 9am the next day, the distribution captures the probability that CLR, FEW, SCT, BKN, and OVC are actually observed at 9am. We compute this distribution for each hour of the next day separately. Since the day-ahead market clears at the same time each day, e.g., 8-10am, each hour of the next day represents a different forecast time horizon, and, in general, the longer the time horizon, the lower the forecast accuracy. We expect weather forecasts later the next day to be less accurate on average than forecasts near the beginning of the day.

Fig. 1 shows the conditional distribution of observed cloud cover readings for a given forecast (in this case, for the next day at 3pm) at a solar site location in Massachusetts. The x-axis shows the forecasted cloud cover, while the y-axis shows the probability, for the given forecasted value, that a certain cloud cover string was observed the next day (at 3pm). This distribution is derived from all cloud cover forecast and observational data over a six year period (2011-2018 excluding 2016) for that location. The graph includes more than 30,000 hourly cloud cover observations (and corresponding day-ahead cloud cover forecasts) covering all daylight hours. As the graph shows, extreme conditions are more predictable than intermediate conditions. Clear sky and overcast day-ahead forecasts are highly accurate with a probability greater than 80% and 75%, respectively, of them being correct. In contrast, forecasts for intermediate conditions are less accurate, since they are inherently more variable than the extremes. That is, clear sky and overcast days tend to be sunny and cloudy, respectively, throughout, while intermediate conditions often transition more frequently between different states of cloudiness (as clouds pass by). Even so, the conditional probability distribution follows the expected trend, as cloudier conditions increase in probability with an increasingly cloudy forecast.

The forecast distribution above is for a single site and different sites will yield different distributions based on their local climate, since some sites experience more predictable weather. To illustrate this observation, we used ground truth observational data to plot the probability of observing a certain type of cloud cover condition at three different locations. We compute the probability below from observational data, where Z is one of the five cloud cover conditions.

$$\Pr(\text{observation} = Z) \quad (2)$$

Fig. 2 shows the probability distribution of measured weather at the 3 locations in the United States: Massachusetts (the same site as above), Colorado, and Texas. Massachusetts has high weather variability, such that all cloud cover conditions have similar probability with overcast and broken clouds occurring >40% of the time. In contrast, Texas has a high probability of clear weather, while Colorado is in the middle with a high fraction of both clear and overcast periods. Interestingly, for all three sites we see that extreme conditions—clear and overcast—have the highest probability of occurrence. The graph indicates that cloud cover varies by location, which necessitates computing our forecast

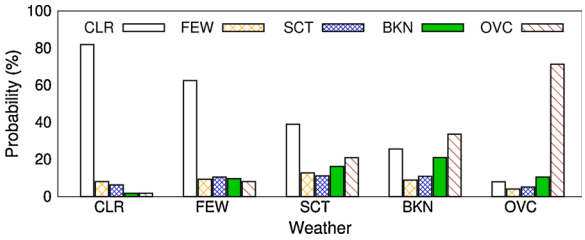


Fig. 1. Example of a site's conditional forecast error distribution, which captures the probability of observing a certain cloud cover on the y-axis given the forecasted cloud cover at 3pm on the x-axis. Data covers all forecasts from 2011-2018 excluding 2016.

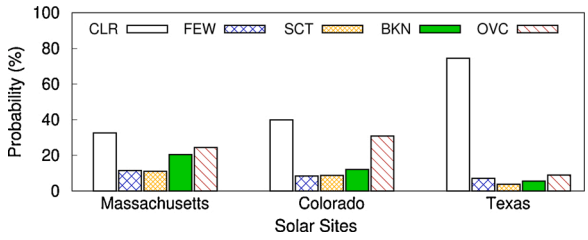


Fig. 2. Observed cloud cover at three locations in the United States with different climates. Data covers 2013-2018.

distributions separately using historical weather data from each location.

Cloud Cover Measurement Uncertainty. We next compute the conditional probability distribution of observing a particular clear sky PV index (K_{pv}) for a given measured cloud cover. As noted earlier, the same measured cloud cover can yield different K_{pv}'s depending on cloud thickness, height, and position in the sky relative to the Sun, which cloud cover measurements do not take into account. We capture this uncertainty, and its resulting effect on solar generation, using a conditional distribution that computes the probability of observing a specific K_{pv} given a cloud cover measurement. That is, we compute the distribution below, where X is K_{pv} and Z is one of the five cloud cover conditions we measure.

$$\Pr(K_{pv} = X | \text{observation} = Z) \tag{3}$$

Our analysis shows that even when grouping cloud cover measurements into coarse levels that each include a wide range of measured cloud cover percentages, there is still significant inaccuracy. This observation is borne out by our empirical data analysis, which shows that the expected K_{pv} varies significantly for the same cloud cover measurement. This demonstrates the current inherent inaccuracy in measuring cloud cover as a coarse percentage at a single location, but then applying it to larger regions.

Fig. 3 shows the conditional probability distribution of measurement error in K_{pv}, i.e., the fraction of solar output relative to a site's maximum output under clear skies, for different observed cloud conditions for the same site as in Fig. 1. The figure shows the measured cloud cover on the x-axis and the probability K_{pv} falls in the specified range on the y-axis for the given measurement. We define the ranges for K_{pv} to be the same as the defined ranges for the different cloud cover conditions. Our analysis uses hourly cloud cover observations and solar generation over four years (2013-2018 excluding 2016). The graph shows that when the sky is clear, the site generates greater than 93% of its maximum clear sky solar output nearly 100% of the time, e.g., K_{pv}>0.93. Similarly, with overcast skies, the probability that solar generation is 0-5% of its maximum output is near 80%. As before, the intermediate cloud covers, especially scattered and broken clouds, are less clear with a mix of probabilities.

Unlike the forecast error distribution which is specific to a site, we can derive our measurement error distribution using data from many sites, since measurement error depends on cloud conditions and not on the location. In particular, we leverage data from prior work that includes 343 million hourly cloud cover and solar readings from over 11,000 locations [20]. We compute this error distribution based on the difference between K_{pv} observations at the solar sites and the estimate from Solar-TK's performance model.

Finally, Fig. 4 illustrates the effect of cloud cover measurement error on solar generation modeling by plotting solar generation for a particular site every hour over a day. The graph shows the actual solar generation (the bottom red curve), the maximum solar generation (the top blue curve), and the normalized solar output K_{pv} (the middle black curve). On this day, while the cloud cover observations were BKN (broken clouds) throughout the day, the graph shows that the actual solar generation varied substantially between nearly 40% of the maximum generation for one hour in the middle of the day to less than 10% at other hours. If the cloud cover measurement were precise and constant throughout the day, we would instead expect the normalized solar output K_{pv} to also be the same fraction throughout the day. The errors stem from the coarseness of the cloud cover measurement and the stochasticity of cloud cover, as clouds pass by.

3.2. Computing the Solar Energy Commitment

We use the two conditional probability distributions from Equations (1) and (3) to compute the solar energy commitment that maximizes a site's expected revenue. We first combine the forecast and measurement error distributions above into a single *conditional probability distribution* that quantifies the probability a solar site will generate a certain K_{pv} for a forecasted cloud cover. The conditional distribution $\Pr(K_{pv} = X | \text{forecast} = Y)$ is computed as shown below [21]. We compute this distribution separately for each hour of the next day.

$$\Pr(K_{pv} = X | \text{forecast} = Y) = \sum_{Z} [\Pr(K_{pv} = X | \text{observed} = Z) \times \Pr(\text{observed} = Z | \text{forecast} = Y)] \tag{4}$$

Here, X takes output values from $0-1$, while Y and Z take on one of the five cloud cover conditions. When computing this conditional probability distribution, we assume that any percentage cloud cover within the percentage okta range defined by the cloud cover reading is equally likely. For example, since the defined okta range for OVC is 7/8 to 8/8 (or 87.5% to 100%), we assume any value between 87.5% and 100% is equally likely. Given this assumption, we convert our coarse forecast error distribution into one that conveys the probability of a specific percentage cloud cover occurring (from 0%-100%) by drawing a random percentage value within the corresponding range for each forecasted cloud cover. Thus, for any OVC forecast, we would select a random percentage between 87.5% and 100%. Doing so, places the forecast distribution on the same scale as the cloud cover measurement error distribution.

The distribution captures the probability a certain K_{pv} occurs based on a forecasted cloud cover at a specific hour. We use this conditional distribution to derive a cumulative distribution function (CDF) of

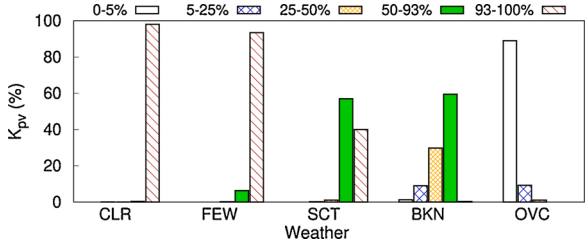


Fig. 3. Example of a site's K_{pv} probability distribution showing the measured cloud cover on the x-axis, and the probability of different K_{pv}'s on the y-axis. Data covers 2013-2018 excluding 2016.

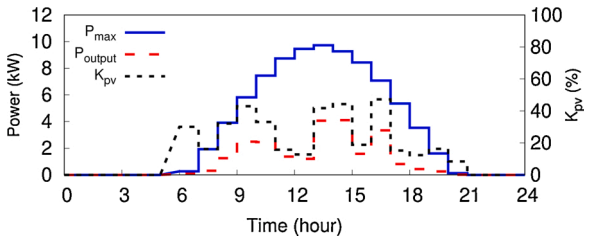


Fig. 4. A site's maximum solar output under clear skies, its actual solar output, and its normalized solar output K_{pv}, i.e., actual output divided by maximum output, on a day when broken clouds (BKN) were reported throughout the day.

possible solar output values K_{pv} for each forecasted hour of the next day, which we denote as $Z(x)$. Fig. 5 shows an example CDF $Z(x)$ for a forecasted hour of cloud cover. The CDF quantifies the probability that the K_{pv} is less than the K_{pv} on the x-axis.

Given the CDF, to compute the solar energy commitment that maximizes expected revenue, we take each point on the x-axis of the CDF and assume it is our solar energy commitment, which we denote as x. We then calculate the expected revenue for committing this x for two cases: one given that our solar output is less than x and another one given that our solar output is greater than x. Each case results in a different revenue equation, as the former loses revenue by not committing enough solar energy, while the latter incurs a deviation penalty for committing too much solar energy.

To compute expected revenue $R(x)$ for a given solar commitment x, we multiply the expected revenue in each case with the probability that each case occurs, as shown below.

$$E[R(x)] = E[R(x)|K_{pv} \leq x] \times P(K_{pv} \leq x) + E[R(x)|K_{pv} > x] \times P(K_{pv} > x) \quad (5)$$

The probability $P(K_{pv} \leq x)$ and $P(K_{pv} > x)$ comes directly from the CDF. We discuss deriving $E[R(x)|K_{pv} \leq x]$ and $E[R(x)|K_{pv} > x]$ in the two cases below.

Over-commitment Case: $K_{pv} \leq x$. In this case, we assume we over-committed solar energy by committing an x that is greater than the K_{pv} we actually generated. We denote $L = E[K_{pv}|K_{pv} \leq x]$, or the expected value of K_{pv} given that it is less than the commitment x. We then compute the expected revenue in this case as shown below.

$$E[R(x)|K_{pv} \leq x] = (L \times \text{price} \times P_{max}) - ((x - L) \times \text{penalty} \times P_{max}) \quad (6)$$

The first term represents the expected earnings from committing L at a certain day-ahead price. Note that we have to multiply by P_{max} (the maximum clear sky solar output), since L (and K_{pv}) are in terms of the fraction of maximum solar output under clear skies, while price is in dollars per absolute kilowatt-hour generated. The second term represents the expected penalty we must pay for over-committing solar energy: it is the difference between the commitment x and expected generation L multiplied by some penalty price per kilowatt-hour of energy. As before, we must multiply by P_{max} to de-normalize the commitment fraction x. Our expected revenue when we over-commit solar energy is then the combination of these terms, i.e., our earnings from selling the energy we generated minus the penalty from over-committing.

Under-commitment Case: $K_{pv} > x$. In this case, we assume we under-committed energy by committing an x that is less than the K_{pv} we actually generated. We denote $H = E[K_{pv}|K_{pv} > x]$, or the expected value of K_{pv} given that it is greater than the commitment x. We compute the expected revenue in this case as shown below.

$$E[R(x)|K_{pv} > x] = \text{price} \times x \times P_{max} \quad (7)$$

Since we under-commit solar energy, our expected revenue is simply the price multiplied by our commitment x multiplied by P_{max} (to de-normalize the commitment fraction x). We do not pay a penalty, since we satisfy our commitment. Note that, while it is not represented in the equation above, we lose revenue based on the difference between H and x, since we could have generated more revenue by increasing our commitment x (without paying a penalty).

Finally, we compute this expected revenue $E[R(x)]$ for each point on the CDF's x-axis and commit the point that maximizes revenue.

$$\underset{x}{\text{maximize}} R(x)$$

We derive a separate CDF from the conditional distribution for each forecasted hour of the next day, and perform the method above, to determine the solar energy commitment each hour that maximizes expected revenue. Fig. 6 summarizes the steps for our probabilistic solar commitment approach.

4. Implementation

Our implementation leverages Solar-TK [3], an open-source solar performance model, to estimate a site's solar energy output based on its location, time, physical characteristics, cloud cover, and temperature. However, our approach is independent of any specific solar performance model, assuming it takes cloud cover as an input. We compute the forecast and measurement error probability distributions using archival forecast data from the NDFD [11], and weather data from the NWS. These sources report hourly cloud cover and temperature forecasts and measurements for many years.

We implement our commitment strategy in Python and use NumPy and Pandas for weather and energy data processing. When computing the forecast and measurement probability distributions we exclude the

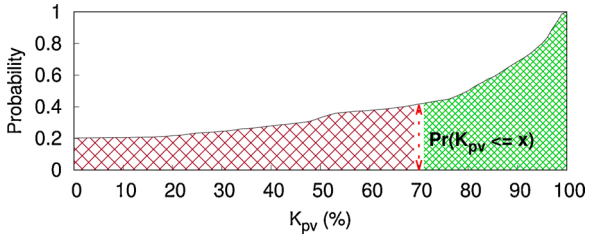


Fig. 5. CDF of K_{pv} for next-day hour based on conditional distribution of cloud cover forecast and measurement error.

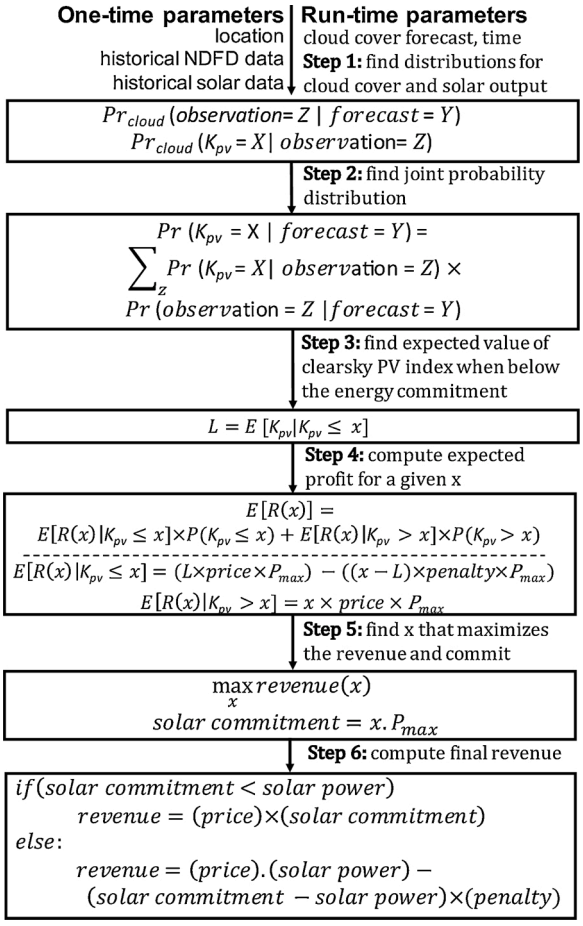


Fig. 6. Flow of steps in our probabilistic approach for deriving a solar energy commitment using the next-day forecast.

first and last hour of sunlight as solar generation during these hours is low and highly variable. We also exclude days when it has snowed, as snow significantly degrades solar output, which can result in additional errors. We assume solar operators commit no solar energy when solar modules are covered by snow. While our probability distributions draw from 5-7 years of data, we evaluate our commitment strategies over one year (2016) using solar generation, weather, and forecast traces from that year. We use 2016 for evaluation as solar data for 2016 had lowest percentage of missing values and incorrect observations. We exclude 2016 data when generating probability distributions. We use real day-ahead electricity prices during this time period for some of our experiments, which are publicly released and archived by ISOs.

We plan to publicly release our implementation as open-source. The software automatically downloads the location's NDFD forecasts and generates the CDF in Fig. 5, and then accesses real-time electricity prices from regional ISO to determine next-day solar commitments using our probabilistic approach from §3.

5. Evaluation

We evaluate our approach by comparing the revenue it produces with multiple other solar commitment strategies, as well as the theoretical optimal. In addition to total revenue, we also evaluate these approaches based on the total solar energy they over- and under-commit. Over-commitments impact grid stability, especially if solar generators over-commit *en masse*, while under-commitments result in a loss of revenue. Since our results are a function of day-ahead and penalty prices, we first discuss their characteristics.

5.1. Electricity Price Characteristics

The magnitude, pattern, and predictability of day-ahead and penalty prices affect the revenue of any solar commitment strategy, including our probabilistic approach, which requires them as input (see Equations (6) and (7)). The characteristics of these prices in the market differ substantially by location and over time. However, in this paper, our goal is to isolate and quantify the effect of inaccurate cloud cover measurements and forecasts on solar commitments and revenue independent of any specific set of price characteristics. As a result, our evaluation covers a wide range of price characteristics.

In the simplest case, we experiment with a fixed day-ahead price (equal to the average day-ahead price of $40.7/MWh at our location over 2016) and a range of fixed penalty prices. We also experiment with variable day-ahead prices (based on the real day-ahead prices over 2016) for a range of fixed penalty prices. We assume accurate day-ahead price forecasts, since our focus is on the impact of inaccurate cloud cover measurements and forecasts. In addition, while day-ahead prices are not known *a priori*, they are highly predictable in the current market, and vary primarily with slow and predictable changes in temperature that affect the heating and cooling load [22].

We evaluate our approach over a wide range of fixed penalty prices to understand how the magnitude of the penalty relative to day-ahead prices affects solar's revenue. Specifically, our penalties range from 0% of the day-ahead price to 250% of the day-ahead price. These penalty levels are consistent with prior work on solar commitment which uses 150% [23] and 200% [24]. Our results confirm this prior work which suggests that penalties must reach ∼200% to strongly incentive accurate predictions [23,24]. Note that lower penalties or the absence of penalties are subject to gaming, and encourages over-committing energy. We expect our probabilistic approach to improve relative to others as the penalty price increases relative to the day-ahead price, since this incentivizes making more accurate commitments.

5.2. Baseline Solar Commitment Strategies

We compare with multiple baseline commitment strategies, including two deterministic solar commitment strategies, as well as the optimal strategy, which assumes perfect solar forecasts.

Max Solar. The *Max Solar* commitment strategy commits a solar site's maximum expected solar generation under clear skies ($P_{\max}$) estimated by Solar-TK each hour of the next day. This maximum clear sky generation varies as a function of location, time, physical characteristics, and the temperature. This approach assumes clear skies at all times. We use this strategy to demonstrate the impact of over-committing solar on revenue, as it always over-commits.

Trust Forecast. The *Trust Forecast* commitment strategy assumes cloud cover measurements and forecasts are correct and forecasts solar energy each hour of the next day using Solar-TK directly. For this approach, we use the fine-grained cloud cover percentage from 0-100% provided by the NDFD. We translate the cloud cover forecast readings into a corresponding expected K_{pv} due to the cloud clover. Prior work has proposed the simple empirically-derived function below for doing this based on an analysis of weather and solar data across thousands of sites [14,25].

$$K_{\text{pv}} = (0.985 - 0.984n^{3.4}) \tag{8}$$

Here, n represents the cloud cover reading from 0.0-1.0. Note that these forecasts and the model used to convert cloud cover into power represent the state-of-the-art in solar power forecasting [26,27,14].

This strategy is deterministic in that it directly translates cloud cover forecasts into solar energy forecasts. The approach does not consider the deviation penalty as it trusts forecasts are correct.

Optimal. The *Optimal* commitment strategy has perfect knowledge of future solar generation. This approach maximizes revenue, as it always commits the maximum possible energy and never incurs a penalty. Since the optimal represents the best we could expect to do, we normalize all our results with respect to optimal.

5.3. Revenue Comparison

We compare our probabilistic approach with the baseline approaches above in terms of the total revenue they generate. Our evaluation uses generation data from the Massachusetts solar site shown in Fig. 2, which has highly variable cloud cover and temperature variations throughout the year. Note that this site is more challenging than the other sites in Fig. 2, since it experiences the intermediate cloud cover conditions for more time (FEW, SCT, and BKN). The solar site has a 10kW capacity and generates ∼11.1MWh of energy per year, which corresponds to a large residential deployment.

Fig. 7 compares the revenue on the y-axis for our different solar commitment strategies as the fixed deviation penalty price increases on the x-axis. This experiment uses a fixed day-ahead price equal to $40.7/MWh. We normalize revenue on the y-axis with respect to the revenue earned by the optimal strategy, and we normalize the penalty price with respect to the average day-ahead price. Thus, a 100% penalty translates to a deviation penalty price of $40.7/MWh, which represents the point where the penalty for over-committing a kilowatt-hour of solar energy is equal to the revenue earned from committing and delivering a kilowatt-hour in the day-ahead market. As expected, the graph shows that with a 0% penalty, all approaches produce at or near the optimal revenue. Even though *Max Solar* always over-commits solar energy, when it incurs no penalty, it produces 100% of the optimal revenue. Similarly, since our *Probabilistic* approach takes into account the penalty, it determines that over-committing is the optimal strategy and also achieves 100% of the optimal revenue. However, since *Trust Forecast* does not consider the penalty when determining commitments, it sometimes under-commits solar energy (when the forecast is cloudier than in reality) and thus only achieves ∼90% of the optimal revenue.

As the penalty increases on the x-axis, *Max Solar*'s revenue rapidly decreases, since it blindly over-commits by assuming cloudless skies and incurs increasingly larger penalties. At a 100% penalty (where the

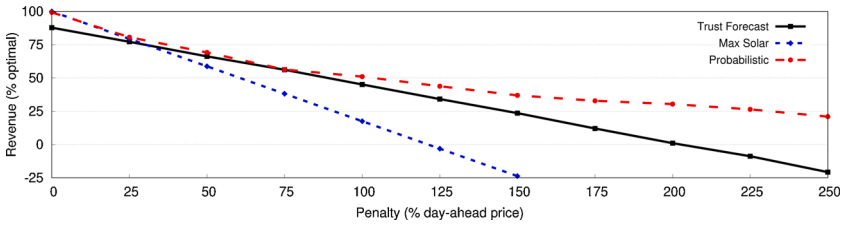


Fig. 7. Revenue for each solar commitment strategy for fixed day-ahead price of $40.7/MWh as the fixed penalty price increases.

penalty and day-ahead price are equal), *Max Solar* generates only ∼20% of the optimal revenue. With a 250% penalty, *Max Solar* generates no revenue but instead incurs a net penalty equal to 25% of the optimal revenue. In contrast, the other approaches are more robust to higher penalties with our *Probabilistic* approach being the most robust. At 100% penalty, the *Probabilistic* approach generates the most revenue at 49% of optimal, while *Trust Forecast* generates nearly 10% less revenue (at ∼45% of optimal). The advantage of the *Probabilistic* approach continues to increase further as the penalty rises. At a 200% penalty, the *Probabilistic* approach still generates over 25% of the optimal revenue, while the other approaches generate no revenue and start to incur a net penalty.

Fig. 8 shows the same experiment, but using real day-ahead prices from 2016, rather than fixed day-ahead prices, where the average day-ahead price is the same in both cases ($40.7/MWh). The graph exhibits the same trends as before, with only very slight differences in the magnitude of revenue. This experiment shows that the variability in the day-ahead price does not have a significant impact on revenue. As a result, our probabilistic approach is not biased to under-committing when day-ahead prices are high, and over-committing when day-ahead prices are low. Both cases would decrease our revenue relative to a fixed day-ahead price.

Figs. 7 and 8 quantify the discounted (or risk-adjusted) value of solar energy in the market (for different deviation penalties) due to its lack of predictability. Since solar generators cannot dispatch or predict the energy they will generate, their commitments impose some risk, which reduces the revenue they earn from the energy they ultimately generate. Based on Figs. 7 and 8, this solar discount is equal to 49% and 47.3% of the optimal value, respectively, when the penalty price increases to be equal to the average day-ahead price. Of course, solar may be worth more in locations with more predictable weather patterns. An important motivation for our approach is to encourage better solar forecasting methods that can increase the value of solar in the market, which can also enable less solar regulation and a higher grid solar penetration.

Our probabilistic approach leverages both cloud cover forecast error distribution and cloud cover measurement error distribution. In our analysis, we observe the effect of both error distributions on different metrics individually. Fig. 7 shows only the combined effect of both distributions. At 100% penalty, accounting for only forecast error generates 46.8% (roughly 4% more revenue than the *Trust Forecast* approach). In energy commitment, accounting for only the forecast error shows the same trend as the complete probabilistic approach and undercommits energy. Finally, the probabilistic approach with forecast error results in 58% frequency of overcommitment (compared with 71% for *Trust Forecast*). It also yields a MAPE of 33%, 13% lower than *Trust Forecast* approach. Our analysis shows that while an approach that accounts for only forecast error improves performance, a probabilistic approach that accounts for both forecast and measurement error has a clear advantage.

In addition, our approach compares favorably to prior work on the solar commitment problem that uses deterministic forecasts [23]. Specifically, this work produced 59% of the optimal revenue at a 150% penalty (averaged across 63 sites in California) with much more generous market assumptions that enabled sites to earn additional revenue by selling surplus solar energy (from under-commitments) in the real-time market. In contrast, our approach, which does not include additional revenue from surplus solar, and produces 42% of the optimal revenue at the same penalty level for a site in Massachusetts with much more variable (and thus difficult to predict) weather. As we show in the next section, our probabilistic approach becomes more conservative as the penalty increases, which results in a high under-commitment levels. For example, we show that at 150% penalty, we under-commit nearly 50% of the energy produced, which, if sold, would significantly increase revenue.

5.4. Energy Commitment Comparison

We also evaluate the the amount of energy each approach over- and under-commits as the penalty price increases. In general, the grid prefers under-commitments rather than over-commitments, since solar sites that under-commit satisfy their commitments. Numerous and frequent over-commitments can lead to price volatility in the spot market (where generators must make up their deficit in over-commitment) and grid instability (if there is not enough supply in the spot market to meet the demand). This graph uses the same variable day-ahead prices as in Fig. 8. For these experiments, we omit *Optimal*, since it never over- or under-commits.

Fig. 9(a) shows the amount of excess energy over-committed as a percentage of the total solar energy generated during over-commitment periods. Since *Max Solar* and *Trust Forecast* do not vary based on the penalty price, their percentage over-commitment is the same in all cases. However, as expected, our probabilistic approach becomes more conservative (and over-commits less) as the penalty increases. With a low penalty it over-commits closer to *Max Solar*, but with a high penalty it over-commits the least, which promotes less price volatility and a more stable grid. Similarly, Fig. 9(b) shows the amount of deficit energy under-committed as a percentage of the total solar energy generated during under-commitment periods. The trend is the opposite of (a) with

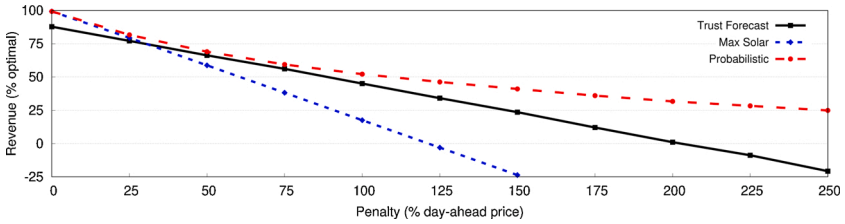


Fig. 8. Revenue for each solar commitment strategy for variable day-ahead prices as the fixed penalty price increases.

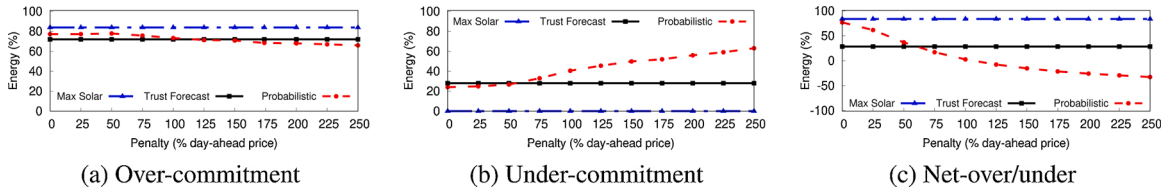


Fig. 9. Fraction of solar energy each policy over-commits (a), under-commits (b), and net over/under-commits (c).

the percentage of energy under-committed increasing as the penalty price increases. Fig. 9(c) shows the balance of the two, with the amount of energy over-committed and under-committed being roughly equal (i.e., crossing x=0) at a 100% penalty.

Finally, Fig. 10 plots i) the Mean Absolute Percentage Error (MAPE) between our day-ahead solar commitments and our actual generation and ii) the percentage of periods where our strategy over-commits. This graph is for a 100% deviation penalty (equal to the day-ahead price), and shows that our probabilistic approach has the most accurate solar energy forecasts (higher than *Trust Forecast*) due to its lower MAPE *and* over-commits with the lowest frequency.

5.5. Increasing Revenue using Energy Storage

While our probabilistic approach produces more revenue than the other approaches as the penalty increases, its revenue still decreases due to inherent inaccuracy in cloud cover measurements and forecasts. In addition, at high penalty prices, our approach loses a significant amount of revenue from under-committing a large fraction of its energy generation. Energy storage is a potentially useful tool in making up the difference between our committed solar energy and our actual solar generation. Energy storage both enables sites to store excess solar energy when they under-commit energy in the market, and release stored energy to make up the deficit when they over-commit energy in the market. To better understand the benefits of energy storage, we implement a policy that stores any excess solar not committed when using our probabilistic approach (when the energy storage is not full), and uses any stored energy (until the stored energy runs out) to make up for deficits when over-committing.

Fig. 11 plots the usable battery capacity on the x-axis and the percentage of the optimal solar revenue when using a battery of that capacity on the y-axis for two different penalty prices (100% and 200% of the day-ahead price). The graph incorporates battery costs based on the cost of the Tesla Powerwall 2, which costs $6500 and has a capacity of 13.5kWh. We assume a battery lifetime of 7 years and amortize the cost over its lifetime. As expected, the battery capacity enables us to partially correct for inaccuracies in cloud cover forecasts and measurements, which has a greater impact at high penalties. The graph shows that the revenue increases as we add more battery capacity up to a point, after which the battery costs being to dominate and the revenue increases. Even so, the graph shows that battery-based energy storage can result in significant revenue increases. Interestingly, for this residential solar site, the optimal battery capacity is near the 13.5kWh capacity of the Tesla

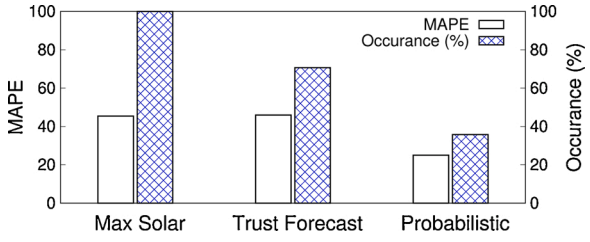


Fig. 10. MAPE between day-ahead solar energy commitments and solar generation with our probabilistic approach, and the percentage of periods where our strategy over-commits.

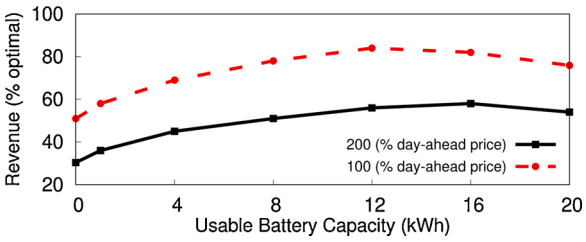


Fig. 11. Revenue on the y-axis from using the battery-based energy storage capacity on the x-axis.

Powerwall 2. Specifically, in the case of a 100% penalty price, the optimal capacity is 12kWh, which yields over 80% of the optimal revenue or a 60% increase compared to not using a battery. Similarly, in the case of a 200% penalty price, the optimal capacity is 16kWh and yields 50% of the optimal revenue, which is 2× more compared to not using a battery.

Note that, in this example, we only use the battery capacity to make up for inaccuracies in our solar energy commitments. As a result, the presence of energy storage would not alter the revenue of the optimal strategy that assumes perfect solar forecasts. Our results above show that, even after incorporating their cost, batteries can make up a significant portion of the lost revenue due to inaccurate solar forecasting. More generally, our work shows that, with a battery, even small-scale residential solar sites could compete in the market and retain much of the revenue they receive today without requiring arbitrary restrictions on grid-tied solar connections. The revenue can potentially be increased further by incorporating the batteries in the solar commitment problem. The availability of storage enables the commitment approach to favor undercommitment or overcommitment based on the stored energy and the expected energy price. However, in this work, we want to isolate the effect of probabilistic forecast approach from the effect of storage and leave the analysis of commitment approach with storage to the future work. In addition, batteries may also provide other valuable services to users, including as a source of backup power or for grid peak shaving. For example, Massachusetts recently introduced a program for compensating residential batteries owners in exchange for using them to reduce grid peaks a few times a year [28]. We defer a full cost-benefit analysis of batteries in the context of these programs to future work.

6. Related Work

Our work is related to prior work on solar forecasting [12]. However, we do not propose a new forecasting technique, but instead examine the impact of inaccurate forecasts on solar commitments. We use an open-source forecasting model to derive forecast and measurement error distributions using publicly-available forecast and weather data. We could apply our techniques to any solar forecasting model. Our underlying model for generating a solar energy forecast from weather and cloud cover data is also similar to prior work [14] and represents the state-of-the-art in terms of both input data and solar modeling, both of which are publicly available.

There has also been significant prior work on bidding strategies for wind and solar energy in electricity markets based on both forecast and

price uncertainties. For example, Jiang et al. provide a survey of approaches for integrating solar into energy markets, and the potential impact of solar on the market [29]. Prior work includes proposals for bidding strategies for wind [30–33], solar [34,35], or both [36,37]. The precise problems these approaches focus on are different. In some cases, the work assumes the use of energy storage, while others assume hybrid solar/wind systems. Similarly, some work focuses on the variations and predictability in energy prices. In many cases, the strategies are evaluated on data drawn from idealized distributions that do not take into account the coarse granularities of measurement and their imprecision. Our approach instead focuses on a more specific problem by focusing only on the uncertainty due to cloud cover forecasts and measurements. We show that considering both results in greater revenue.

7. Conclusions

Currently, complex regulations govern who can connect solar to the grid, and how much compensation they receive for it. Ultimately, however, for solar to expand, it must compete on equal footing with other energy sources. Thus, an alternative approach is to have solar generators compete with traditional sources in existing energy markets. However, a key problem for solar operators is determining how much energy to commit in day-ahead markets, especially given uncertain forecast and cloud cover measurements. We address this problem by developing a probabilistic approach to forecasting and committing solar energy in day-ahead markets that accounts for the uncertainty in cloud cover measurements and forecasts.

Our approach determines a joint probability distribution over next-day solar generation outcomes, which we then use to determine solar energy commitments each hour that maximize expected revenue. We show that, as the penalty for over-committing solar increases, our probabilistic approach enables increasingly more savings than a deterministic approach that trusts cloud cover measurements and forecasts. Since solar's unpredictability decreases its revenue in the market, we then evaluate the impact of using energy storage to recoup the lost revenue. We show that batteries can make up a significant fraction of lost revenue, and enable even small-scale solar sites to effectively compete in the market.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgement

This research was supported by NSF grant #1645952.

References

[1] J. Blatchford, Participating Intermittent Resource Program (PIRP), CAISO.
[2] Real-Time vs. Day-Ahead Pricing, 2018 https://www.aepenergy.com/2018/01/05/december-2017-edition/.
[3] N. Bashir, D. Chen, D. Irwin, P. Shenoy, Solar-TK: A Data-driven Toolkit for Solar PV Performance Modeling and Forecasting, MASS (2019).
[4] V. Arya, P. Dutta, S. Kalyanaraman, On Mitigating Wind Energy Variability with Storage, COMSNETS (2013).
[5] Overview of New England's Wholesale Electricity Markets and Market Oversight, 2014 https://www.iso-ne.com/pubs/spcl_rpts/2014/2014_market_overview_050614.pdf.
[6] A Guide to Electricity Markets, Systems, and Policy in Massachusetts, 2015 https://www.mountgrace.org/sites/default/files/Electricity-Markets-Primer.CLF_.pdf.
[7] Weather Research and Forecasting Model Users Page, 2019 http://www2.mmm.ucar.edu/wrf/users/.
[8] Weather.gov Terms, http://www.weather.gov/bgm/forecast_terms, ????.
[9] Automated Surface Observing System (ASOS) User's Guide, 2018 https://www.weather.gov/media/asos/aum-toc.pdf.
[10] M. Weiss, J. Ghirardelli, A Summary of Ceiling Height and Total Sky Cover Short-Term Statistical Forecasts in the Localized Aviation MOS Program (LAMP), AMS Conference on Weather Analysis and Forecasting (2005).
[11] National Digital Forecast Database (NDFD), 2018. URL https://www.weather.gov/mdl/ndfd_home.
[12] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.M. de Pison, F. Antonanzas-Torres, Review of Photovoltaic Power Forecasting, Solar Energy 136.
[13] E. Lorenz, D. Heinemann, H. Wickramarathne, H. Beyer, S. Bofinger, Forecast of Ensemble Power Production by Grid-connected PV Systems, 20th European PV Conference, Milano, 3-9 (2007).
[14] R. Perez, K. Moore, S. Wilcox, D. Renné, A. Zelenka, Forecasting Solar Radiation - Preliminary Evaluation of an Approach Based Upon the National Forecast Database, Solar Energy 81 (6) (2007) 809–812. ISSN 0038-092X.
[15] N. Bashir, D. Irwin, P. Shenoy, Helios: A Programmable Software-defined Solar Module, BuildSys (2018).
[16] D. Chen, D. Irwin, Black-box Solar Performance Modeling: Comparing Physical, Machine Learning, and Hybrid Approaches, Greenmetrics (2017).
[17] N. Engerer, F. Mills, Kpv: A Clear-sky Index for Photovoltaics, Solar Energy 105.
[18] Integrated Surface Database (ISD), 2018. URL https://www.ncdc.noaa.gov/isd.
[19] Weather Underground, 2018. URL https://www.wunderground.com/history/.
[20] D. Chen, J. Breda, D. Irwin, Staring at the Sun: A Physical Black-box Solar Performance Model, BuildSys (2018).
[21] A. Papoulis, S. Pillai, Probability, Random Variables and Stochastic Processes, chap. 2, Tata McGraw-Hill Education, 2002.
[22] C. Rodriguez, G. Anders, Energy Price Forecasting in the Ontario Competitive Power System Market, IEEE Transactions on Power Systems 19 (1) (2004) 366–374.
[23] J. Luoma, P. Mathiesen, J. Kleissl, Forecast Value Considering Energy Pricing in California, Applied energy 125 (2014) 230–237.
[24] J. Kleissl, Solar Energy Forecasting and Resource Assessment, Academic Press, 2013.
[25] D. Chen, J. Breda, D. Irwin, Staring at the Sun: A Physical Black-box Solar Performance Model, Proceedings of the 5th Conference on Systems for Built Environments (2018) 53–62.
[26] R. Perez, S. Kivalov, J. Schlemmer, K.H Jr., D.R.T.E. Hoff, Validation of Short and Medium Term Operational Solar Radiation Forecasts in the US, Solar Energy 84 (12) (2010) 2161–2172.
[27] R. Perez, J. Schlemmer, S. Kivalov, J. Dise, P. Keelin, M. Grammatico, T. Hoff, A. Tuohy, A New Version of the SUNY Solar Forecast Model: A Scalable Approach to Site-Specific Model Training, 2017 IEEE 45th Photovoltaic Specialists Conference (PVSC) (2018).
[28] Massachusetts Connected Solutions Program, https://www.nationalgridus.com/MA-Home/Connected-Solutions/BatteryProgram, ????.
[29] S. Jiang, C. Wan, C. Chen, E. Cao, Y. Song, Distributed Photovoltaic Generation in the Electricity Market: Status, Mode and Strategy, CSEE Journal of Power and Energy Systems 4 (3) (2018) 263–272.
[30] S. Kurandwad, C. Subramanian, V. Ramakrishna, A. Vasan, V. Sarangan, V. Chellaboina, A. Sivasubramaniam, Windy with a Chance of Profit: Bid Strategy and Analysis for Wind Integration, in: e-Energy (2014).
[31] I. Gomes, H. Pousinho, R. Melico, V. Mendes, Optimal Wind Bidding Strategies in Day-Ahead Markets, in: DoCEIS (2016).
[32] A. Giannitrapani, S. Paoletti, A. Vicino, D. Zarrilli, Bidding Wind Energy Exploiting Wind Speed Forecasts, IEEE Transactions on Power Systems 31 (4) (2016) 2647–2656.
[33] A. Tha.tte, L. Xie, D.E. Viassolo, S. Singh, Risk Measure Based Robust Bidding Strategy for Arbitrage Using a Wind Farm and Energy Storage, IEEE Transactions on Smart Grid 4 (4) (2013) 2191–2199.
[34] R. Hytowitz, K. Hedman, Managing Solar Uncertainty in Microgrid Systems with Stochastic Unit Commitment, Electric Power Systems Research 119 (2015) 111–118.
[35] A. Giannitrapani, S. Paoletti, A. Vicino, D. Zarrilli, Bidding Strategies for Renewable Energy Generation with Non Stationary Statistics, IFAC Proceedings Volumes 47 (3) (2014) 10784–10789.
[36] G. Liu, Y. Xu, K. Tomsovic, Bidding Strategy for Microgrid in Day-Ahead Market Based on Hybrid Stochastic/Robust Optimization, IEEE Transactions on Smart Grid 7 (1) (2016) 227–237.
[37] I. Gomes, H. Pousinho, R. Melico, V. Mendes, Bidding and Optimization Strategies for Wind-PV Systems in Electricity Markets Assisted by CPS, Energy Procedia 106 (2016) 111–121.