Advanced Scenario Creation Strategies for Stochastic Economic Dispatch with Renewables

Ryan N. King, Matthew Reynolds, Devon Sigler, and Wesley Jones

Abstract—Real-time dispatch practices for operating the electric grid in an economic and reliable manner are evolving to accommodate higher levels of renewable energy generation. In particular, stochastic optimization is receiving increased attention as a technique for handling the inherent uncertainty in wind and solar generation. The typical two-stage stochastic optimization formulation relies on a sample average approximation with scenarios representing errors in forecasting renewable energy ramp events. Standard Monte Carlo sampling approaches can result in prohibitively high-dimensional systems for optimization, as well as a poor representation of extreme events that challenge grid reliability. We propose two alternative scenario creation strategies, importance sampling and Bayesian quadrature, that can reduce the estimator’s variance. Their performance is assessed on a week’s worth of 5 minute stochastic economic dispatch decisions for realistic wind and electrical system data. Both strategies yield more economic solutions and improved reliability compared to Monte Carlo sampling, with Bayesian quadrature being less computationally intensive than importance sampling and more economic when considering at least 20 scenarios.

Index Terms—Stochastic programming, economic dispatch, importance sampling, Bayesian quadrature

I. INTRODUCTION

M ODERN practices for economic dispatch of the electrical grid are rooted in deterministic problem formulations. The increasing deployment of stochastic renewable energy generators, such as wind and solar, requires dispatch techniques that are consistent with the reliability and economic operating requirements of the electric grid. Stochastic optimization techniques have received increased attention from the power systems community, and are a promising way to address the randomness in renewable energy generation. This paper explores two novel approaches to scenario-based stochastic optimization that yield lower costs and reduced reserve requirements, facilitating the grid integration and further deployment of renewable energy generators.

As access to high performance computing resources grows, grid operators are able to consider increasingly sophisticated optimization problems in their dispatch decisions. These complexities can include considering AC optimal power flow (ACOPF), larger grids with thousands of buses, multiple stages or time periods, and stochastic optimization techniques. These techniques can improve grid reliability in the face of uncertain generation, and help remove obstacles to further renewable energy deployment.

An emerging approach to dispatching generators in the face of uncertain renewable generation is to formulate and solve a two-stage stochastic programming problem with recourse. In two-stage recourse problems a first stage decision is made, followed by a recourse stage where adjustments can be made after a random variable is realized. In the context of economic dispatch for the grid, this involves making initial generator dispatch decisions that minimize the combined costs of the first stage and the expectation of the recourse costs. The expected recourse costs are calculated with respect to the distribution of the uncertain renewable generation, and are typically estimated with a sample average approximation using Monte Carlo techniques [1], [2]. However, Monte Carlo techniques exhibit slow convergence with respect to the number of samples and draw samples that are not necessarily a good test of grid robustness in the event of large ramps in generation, creating a need for improved sampling techniques.

In this article we report on two techniques for creating scenarios of errors in forecasting wind generation that improve the convergence rates and accuracy when calculating the expected recourse costs, ultimately achieving more economic dispatch solutions. The created scenarios also contain ramp events that are more challenging for the grid, resulting in dispatch decisions that produce more reliable grid operation. The

NOMENCLATURE

\[ x_g \] Dispatch of thermal generators
\[ c_g \] Cost of thermal generator \( g \)
\[ L(\cdot) \] Loss function
\[ c_w \] Cost of wind generator
\[ \omega_w \] Dispatch of wind generator
\[ \omega_{\text{w,fcst}} \] Forecasted wind generation
\[ \xi_w \] Error in wind forecast
\[ x_{\text{max}}^g \] Max output of generator \( g \)
\[ x_{\text{min}}^g \] Min output of generator \( g \)
\[ R_{\text{up}}^g \] Max ramp up of generator \( g \)
\[ R_{\text{down}}^g \] Max ramp down of generator \( g \)
\[ G \] Set of thermal generators
\[ T \] Set of timesteps
\[ W \] Set of wind plants
\[ S \] Set of wind forecast error scenarios
\[ D \] Set of load buses
\[ d_q \] Value of loads
\[ y_{\text{q,pl}} \] Loss-of-load
\[ y_{\text{q,pl}} \] Excess capacity
\[ c_{\text{q,pl}} \] Cost of Loss-of-load
\[ c_{\text{q,pl}} \] Cost of excess capacity
\[ \omega_{\text{spl}} \] Spilled wind
\[ c_{\text{w,pl}} \] Cost of spilled wind

R. N. King, M. Reynolds, D. Sigler, and W. Jones are with the National Renewable Energy Laboratory (NREL), Golden, CO, 80401 USA; e-mail: (ryan.king, matthew.reynolds, devon.sigler, wesley.jones@nrel.gov).
Manuscript received June 6, 2018.
cost savings and reliability improvements are demonstrated in a week-long simulation of 5 minute dispatch decisions on realistic wind and grid data.

In Section II we provide a background on stochastic programming techniques for two-stage recourse problems, the standard sample average approximation, and power grid applications in economic dispatch. In Section III we introduce new scenario creation strategies using importance sampling and Bayesian quadrature. In Section IV we demonstrate the implementation of these techniques in a stochastic economic dispatch framework and examine their performance. Finally, we conclude in Section V with a discussion of the results and extensions to more complex multi-stage problems and more complex network topologies.

II. TWO-STAGE STOCHASTIC PROGRAMMING

A. General Problem Formulation

The general two-stage stochastic programming problem \(\text{[3], [4]}\) is defined as:

\[
\min_x f(x) + \mathbb{E}_\xi [Q(x, \xi)],
\]

where \(x\) are the first stage decision variables, \(Q(x, \xi)\) is the second-stage cost function defined by,

\[
Q(x, \xi) = \min_y q(x, \xi)
\]

s.t. \(T_\xi x + W_\xi y = b_\xi,\)

and \(y\) are the second-stage variables, and \(\xi\) are random variables. General constraint equations are represented by \(T_\xi, W_\xi\) and \(b_\xi\), and will be made specific to grid applications in the next section. We restrict ourselves to using the linear stochastic programming problem, i.e.

\[
\min_x c^T x + \mathbb{E}_\xi [L(x, \xi)],
\]

where the recourse function is defined as the solution to,

\[
L(x, \xi) = \min_y c_\xi^T y
\]

s.t. \(T_\xi x + W_\xi y = b_\xi\)

\(y \geq 0.\)

In the context of economic dispatch, the first stage variables \(x\) represent the dispatch levels of all generators, \(c\) are linear costs, and the expectation \(\mathbb{E}_\xi [\cdot]\) is taken with respect to the random variable \(\xi\), which represents errors in forecasting the available wind generation.

The second stage recourse variable, \(y\), represents adjustments to generation using reserves, such as up/down regulation, flex reserve or automatic generation control (AGC), that are necessary based on the amount of wind generation that materializes.

The optimization problem embedded inside the expectation \(\mathbb{E}_\xi [\cdot]\) makes it difficult to solve, so a sample average approximation (SAA) is commonly used to represent the expectation. In the SAA approach, we introduce a set of \(N\) scenarios, \(\xi = \{\xi_i\}_{i=1}^N\), and approximate the expectation as \(\mathbb{E}_\xi [L(x, \xi)] \approx \frac{1}{N} \sum_{i=1}^N L(x, \xi_i)\). Commuting the summation and minimization, we can simplify to the extensive form of a two-stage stochastic programming problem:

\[
\begin{array}{ll}
\min_{x, y, i=1,...,N} & c^T x + \frac{1}{N} \sum_{i=1}^N c_{\xi_i}^T y_i \\
\text{s.t.} & Ax = b \\
& T_{\xi_i} x + W_{\xi_i} y_1 = b_1 \\
& T_{\xi_i} x + W_{\xi_i} y_2 = b_2 \\
& \vdots \\
& T_{\xi_i} x + W_{\xi_i} y_N = b_N \\
& x \geq 0, \quad y_1 \geq 0, \quad y_2 \geq 0, \quad \ldots, \quad y_N \geq 0.
\end{array}
\]

A major challenge in the SAA approach is selecting the scenarios \(S\) such that the expectation \(\mathbb{E}_\xi [L(x, \xi)]\) is accurately and efficiently calculated, while also ensuring reliable operation of the grid during different levels of wind generation. In Monte Carlo approaches, the samples are drawn from their nominal distribution \(p(\xi)\) and given equal weights. This approach yields \(O(N^{-1/2})\) convergence, requiring prohibitively many samples to achieve an acceptable error level. Other methods for selecting scenarios have been recently proposed, including scenarios motivated by grid reserve requirements \(\text{[5]}\), chance constraints \(\text{[6], [7]}\), polynomial chaos expansions \(\text{[8]}\), or epi-spline basis functions \(\text{[9]}\).

B. Two-Stage Stochastic Economic Dispatch

In this study we restrict ourselves to studying the stochastic economic dispatch problem, i.e. balancing the (stochastic) demand for power across a grid with the (stochastic) power generation from all generators on the network. This study considers aggregated wind and load data, and therefore our model is formulated without network constraints. However, the scenario creation methods presented in this study should generalize to more complicated constraints on the network, e.g. DC and ACOPF. For a review of the economic dispatch problem, as well as the DCOPF and ACOPF problems, we recommend \(\text{[10], [11], [12]}\). For an introduction to stochastic grid operations we recommend \(\text{[13]}\).

In the context of grid operations, the stochastic economic dispatch problem takes the following form:

\[
\begin{array}{ll}
\min_x & \sum_{g \in G} (c_g x_g + \mathbb{E}_\xi [L(x, \xi)]) \\
\text{s.t.} & x_g^{\text{min}} \leq x_g \leq x_g^{\text{max}} \quad \forall g \\
& -R_g^{\text{down}} \leq x_g - I_g \leq R_g^{\text{up}} \quad \forall g,
\end{array}
\]

where \(x = x_g\) for \(g\) in the set of generators, \(G\), are the first stage decisions representing thermal dispatch, and \(L(x, \xi)\) is the recourse function for a particular realization of the random variable \(\xi\), which represents the error in forecasting wind generation at each wind plant \(w\) in the set of wind plants.
where $W$. The recourse function is defined as the solution to the following optimization problem:

$$
\begin{align*}
\min_{y^+, y^-, \omega^+ \omega^-} & \quad L(x, \xi) = \\
& \sum_{w \in W} (c_w \omega_w + c_w^{spl} \omega_w^{spl}) + \\
& \sum_{q \in G} (c_q^+ y_q^+ + c_q^- y_q^-) + \\
\text{s. t.} & \quad 0 \leq y_q^+ \forall q \\
& \quad 0 \leq \omega_w \leq \omega_w^{fcst} + \xi_w \forall w \\
& \quad \omega_w^{spl} = (\omega_w^{fcst} + \xi_w) - (\omega_w) \forall w \\
& \quad \sum_{q \in D} (y_q^+ - y_q^-) + \sum_{w \in W} \omega_w + \sum_{q \in G} x_g = \sum_{q \in D} (d_q),
\end{align*}
$$

(7)

where $\xi = \xi_w$, $\omega = \omega_w$, $\omega^{spl} = \omega_w^{spl}$ for $w \in W$ and $y^\pm = y_q^\pm$ for each bus $q$ in the set of buses $D$. The expectation $E_\xi [\cdot]$ is taken with respect to the random variable $\xi$. Generator ramp constraints are imposed between sequential timesteps, and not allowed to participate in recourse. The loss of load or excess capacity represented by $y^\pm$ is a slack variable that facilitates finding feasible solutions and represents a challenge to grid reliability and incurs substantial costs.

We assume the wind forecast errors are relative to a persistence forecast based on how well this approach performs in forecasting over short time-periods (see e.g. [14]). For an in-depth survey of wind power forecasting methods we recommend [13] to the reader.

III. STRATEGIES FOR SCENARIO GENERATION

A. Importance Sampling

Solving the two-stage stochastic economic dispatch problem requires a method of selecting scenarios. A logical starting point would be to draw from the nominal distribution of forecast errors in the wind power. However, this approach will not take into account costs communicated by the loss function (7), potentially resulting in scenarios that do not include information about expensive events. As an example, consider a single wind farm generating power for one time period. Assuming that the nominal distribution of forecast errors is symmetric about zero, simply sampling from this distribution will produce scenarios that represent positive forecast errors, i.e. more wind than forecast, about half the time. If we use a small number of scenarios, e.g. 5 to 10, as is common in stochastic optimization problems, the scenario set may not include scenarios that represent negative forecast errors. This would be a costly mistake since both loss-of-load and reserves are more expensive than dispatching extra generation during the first stage decision process.

To address this issue with random sampling we employ importance sampling, see e.g. [16]. The reason for using importance sampling is that by including information from the cost function $L(x, \xi)$, we construct a new probability density function that assigns more mass to values of $\xi$ corresponding to errors in overpredicting the available wind generation. Additionally, the construction of the importance distribution also provides a formula for properly weighting these samples to ensure the weighted average actually converges to $E_\xi [L(x, \xi)]$. Finally, importance sampling can also yield a lower variance estimate of the expected recourse cost.

Recall that our expected recourse cost is

$$
\mu = E_\xi [L(x, \xi)] = \int_{\Omega} L(x, \xi) p(\xi) d\xi
$$

where $p(\xi)$ is a probability density function on $\mathbb{R}^d$ and $p(\xi) = 0$ for all $\xi \notin \Omega$. Introducing a new pdf $q$ on $\mathbb{R}^d$ (the importance distribution) we can write

$$
\mu = E_q [L(x, \xi)] = \int_{\Omega} L(x, \xi) p(\xi) d\xi
$$

(9)

$$
= \int_{\Omega} \frac{L(x, \xi) p(\xi)}{q(\xi)} q(\xi) d\xi
$$

(10)

$$
= E_q \left[ \frac{L(x, \xi) p(\xi)}{q(\xi)} \right]
$$

(11)

where $E_q [\cdot]$ denotes expectation with $\xi$ distributed as $q$. The ratio $p(\xi)/q(\xi)$ is an adjusted weight on sample evaluations of $L$ that reflects sampling from $q$ instead of $p$.

The importance sampling estimate of $\mu$ is then

$$
\hat{\mu}_q = \frac{1}{N} \sum_{i=1}^{N} \frac{L(x, \xi_i) p(\xi_i)}{q(\xi_i)},
$$

(12)

where $\xi_i$ are realizations of the random variable $\xi$ with distribution $q(\xi)$. The optimal importance sampling distribution, defined as $q(\xi) = L(x, \xi) p(\xi)/\mu$, is a zero-variance estimator, and computes the correct expectation from a single sample. Such an estimator suffers from the curse of circularity (the quantity being estimated, $\mu$, is required to properly normalize the importance distribution), but provides guidance on constructing an approximate importance sampling distribution.

Our approach to constructing an importance distribution is to solve a deterministic economic dispatch problem to approximate $\mu$ and $L(x, \xi)$, and then use them to build an approximation of $q(\xi)$. Our first simplification involves approximating the loss function as $L(x, 0)$, i.e. we assume that the wind forecast is perfect. This step reduces the two stage stochastic optimization problem to a simple linear program:

$$
\begin{align*}
x^* &= \arg \min_x & \sum_{g \in G} (c_g x_g + L(x, 0)) \\
\text{s. t.} & \quad x_g^{\min} \leq x_g \leq x_g^{\max} \forall g \\
& \quad -P_g^{\text{down}} \leq x_g - I_g \leq P_g^{\text{up}} \forall g,
\end{align*}
$$

(13)

which we solve for $x$. Given the solution $x^*$ to the simplified dispatch problem we compute the expectation $\hat{\mu} = E[L(x^*, \xi)]$ via the trapezoidal rule. We remark that this process is fast since each evaluation of the loss function $L(x^*, \xi)$ is computed by solving a linear program. Given the approximated expected cost, $\hat{\mu}$, the importance sampling distribution is defined as

$$
q(\xi) = \frac{L(x^*, \xi)}{\hat{\mu}} p(\xi),
$$

(14)
and \( q ( \xi ) \) is then sampled instead of \( p ( \xi ) \) in the solution of the original stochastic economic dispatch problem.

### B. Gaussian Process Representation

One shortcoming in the importance sampling approach presented in the preceding sub-section is that in reality, \( x \) is a first-stage variable in a two-stage stochastic optimization problem, and its value is computed at the same time as the solution to the rest of the problem. By ignoring this coupling, our proxy \( x^* \) may be off by a significant amount, leading to an importance distribution with an incorrect shape and location.

We propose an alternative to using the importance sampling approach: representing the loss function as a Gaussian process (GP) and computing its expectation via a Bayesian quadrature technique. One benefit of this approach is that handles uncertainty in the second stage loss function introduced by not knowing the first stage dispatch. In other words, we can treat \( L ( \xi ) \) as a random function when \( x \) is unknown. Additionally, computing expectations of the GP with Bayesian quadrature does not require knowledge of the loss function to select the quadrature points. Instead, the Bayesian quadrature algorithm spreads out the sample points \( \xi_i \) in the domain based on the nominal distribution \( p ( \xi ) \) and the Gaussian process covariance kernel function. The Bayesian quadrature algorithm also learns quadrature weights, thus ensuring that the weighted sum of the loss function sampled at the quadrature points returns the correct value for the expectation loss. A comprehensive review of Gaussian processes is provided by Rasmussen and Williams [17], while a description of the Bayesian quadrature sampling strategy can be found in [18].

By representing the loss function as a GP, we make the assumption that the joint distribution of any finite number of loss function evaluations is a multivariate Gaussian. The GP is fully specified by a mean function \( m ( \xi ) = \mathbb{E} [ L ( \xi ) ] \) and a kernel function \( k ( \xi , \xi' ) = \mathbb{E} [ L ( \xi ) - m ( \xi ) ] ( L ( \xi' ) - m ( \xi' ) ) \). When represented as a GP, the loss function is

\[
L ( \xi ) \sim \mathcal{GP} \left[ m ( \xi ) , k ( \xi , \xi' ) \right],
\]

and returns a Gaussian distribution when evaluated at a particular value \( L ( \xi^* ) \). The kernel function describes the covariance between any two points in the GP and encodes structural properties of the underlying random function like the smoothness, periodicity, and stationarity. The kernel function is a nonlinear measure of similarity between inputs, with the dot product kernel leading to standard linear regression. In this study we use a common nonlinear kernel function, the Gaussian or squared exponential function:

\[
k ( \xi , \xi' ) = \tau^2 \exp \left( -\frac{|\xi - \xi'|^2}{2l^2} \right)
\]

where \( \tau \) describes the magnitude and \( l \) describes the smoothness of the function or correlation length between two points.

The GP can be conditioned on a set of loss function training points \( \Xi = \{ \xi_i \}_{i=1}^N \) to improve the posterior estimate of the loss function at new test points. The joint distribution of training outputs \( L = L ( \Xi ) \) and a test output \( L^* = L ( \xi^* ) \) is

\[
\begin{bmatrix}
L \\
L^*
\end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix}
m ( \Xi ) \\
m ( \xi^*)
\end{bmatrix} , \begin{bmatrix}
k ( \xi , \Xi ) & k ( \xi , \xi^*) \\
k ( \xi^*, \Xi ) & k ( \xi^*, \xi^*)
\end{bmatrix} \right).
\]

(17)

Conditioning on the training data \( D = \{ \Xi , L \} \) yields the following estimate for \( L^* \):

\[
L ( \xi^* | D ) \sim \mathcal{N} \left[ m_D ( \xi^* ) , \text{cov}_D ( L ( \xi^* ) , L ( \xi ) ) \right]
\]

(18)

\[
m_D ( \xi^* ) = k_1^T K^{-1} L
\]

(19)

\[
\text{cov}_D ( L ( \xi^* ) , L ( \xi ) ) = k ( \xi^* , \xi^* ) - k_1^T K^{-1} k
\]

(20)

where we have introduced the shorthand \( k_s = k_{s,i} = k ( \Xi_i , \xi^* ) \) and \( K = K_{ij} = k ( \Xi_i , \Xi_j ) \). Note that the mean prediction \( m_D ( \xi^* ) \) is simply a linear combination of the training points \( L \) and requires inverting a matrix that scales with the number of training points, \( N \). Additionally, the posterior covariance is given by the prior covariance less the information gained from the training points.

### C. Bayesian Quadrature

Recall that in stochastic economic dispatch we are faced with calculating the expectation of our GP loss function \( \mu = \int L ( \xi ) p ( \xi ) d\xi \). In Bayesian quadrature [18], [19], [20] the integral \( \mu \) is modeled as a Gaussian random variable because integration is a linear transformation that preserves Gaussianity and the integrand, \( L ( \xi ) \), is a Gaussian process. The density function \( p ( \xi ) \) must also be a Gaussian, however this can be achieved by re-weighting via importance sampling.

Bayesian quadrature allows us to express the expectation and variance of the loss analytically, and then devise a sampling strategy to optimally reduce this variance. Following the derivation in [18], the loss averaged over possible functions is simply the expectation of the mean GP loss function:

\[
\mathbb{E}_{L|D} [ \mu ] = \int \int L ( \xi ) p ( \xi ) d\xi \ p ( L | D ) dL
\]

(21)

\[
= \int \left[ \int L ( \xi ) p ( L | D ) dL \right] p ( \xi ) d\xi
\]

(22)

\[
= \int m_D ( \xi ) p ( \xi ) d\xi
\]

(23)

\[
= \int k ( \xi , \Xi ) K^{-1} L p ( \xi ) d\xi.
\]

(24)

Similarly, the expected variance of the estimate with respect
to possible functions is
\[
\mathbb{V}_{L|D}[\mu] = \int \left[ \int L(\xi) p(\xi) d\xi - \int m_D(\xi') p(\xi') d\xi' \right]^2 p(L|D) dL
\]
\[
= \int \int \left[ [L(\xi) - m_D(\xi)][L(\xi')] - m_D(\xi') \right] p(L|D) dL p(\xi) p(\xi') d\xi d\xi'
\]
\[
= \int \int \text{cov}_D \left( L(\xi), L(\xi') \right) p(\xi) p(\xi') d\xi' d\xi' \tag{25}
\]
\[
= \int \int \left[ k(\xi, \xi') - k(\xi, \xi) K^{-1} k(\xi, \xi') \right] p(\xi) p(\xi') d\xi' d\xi'. \tag{26}
\]

A useful sampling strategy is to choose the set of samples that minimizes the expected variance:
\[
\Xi^* = \arg\min_{\Xi} \mathbb{V}_{L|D}[\mu] . \tag{29}
\]

This can be compactly represented as
\[
\Xi^* = \arg\min_{\Xi} Z - w^T K^{-1} w , \tag{30}
\]
where we have defined the following quantities
\[
w = w_i := \int k(\xi, \Xi_i) p(\xi) d\xi \tag{31}
\]
\[
Z = \int \int k(\xi, \xi') p(\xi) p(\xi') d\xi d\xi'. \tag{32}
\]

Because the integrated prior covariance $Z$ has no dependence on the sample points, the optimization can be simplified to
\[
\Xi^* = \arg\max_{\Xi} w^T K^{-1} w , \tag{33}
\]
which shows that the optimal sampling strategy only depends on the choice of kernel covariance function $k(\xi, \xi')$ and the nominal distribution $p(\xi)$, but not on the actual loss function evaluations $L(\Xi)$. To develop some intuition for Eq. \ref{eq:optimal_samp}, we can see that maximization involves a balance between selecting points that reflect the nominal distribution $p(\xi)$ while avoiding redundancies from selecting similar points.

That the sampling strategy is independent of the loss function evaluations is a key differentiator between importance sampling and Bayesian quadrature approaches. This is an advantage for Bayesian quadrature in cases where the loss function is unknown or stochastic, e.g., due to uncertainties in electrical loads. Furthermore, as a quadrature rather than sampling strategy, the Bayesian quadrature points can be reused at successive timesteps if $k(\xi, \xi')$ and $p(\xi)$ are unchanged. The Bayesian quadrature estimate of the expected loss function takes the form of the following quadrature rule
\[
\mathbb{E}_{L|\Xi, L}[\mu] \approx w^T K^{-1} L = \sum_{i=1}^{N} \gamma_i L(\Xi_i) \tag{34}
\]

IV. PERFORMANCE OF DIFFERENT SAMPLING STRATEGIES

1) Test data: To study the performance of our economic dispatch methods we examine a flattened version of the Reliability Test System, modified by the Grid Modernization Lab Consortium (RTS-GMLC) \cite{22}. The RTS-GMLC system is an update to the original 1979 \cite{22} and 1996 \cite{23} version of the RTS that contains 73 buses, 96 conventional thermal generators and 4 wind plants. However, for our experiments we flatten the network and only consider power balance, i.e., matching the aggregated power output from the 96 conventional thermal generators and 4 wind plants to the aggregated demand without taking into consideration the network.

The wind plant data used in our experiments is drawn from the Wind Integration National Dataset (WIND) Toolkit \cite{24}. To appropriately model wind farms using WIND Toolkit data, wind plant power outputs were aggregated by first assigning geographic locations to nodes in the RTS-GMLC system and then summing the output power from the wind plants nearest to the node locations until the prescribed annualized capacity for the node in the RTS-GMLC system was reached.

The experiments in this section take place during a week in July. Load data from the RTS-GMLC system was combined with simulated wind data from 2013 from the WIND Toolkit. To determine $p(\xi)$, we fit a Student’s t-distribution to errors in a 5 minute persistence forecast. A t-distribution was chosen because of the non-normal, heavy tailed behavior of changes in wind generation. The observation heavy-tailed distributions provide a good fit to wind data has been noted previously, see, e.g., \cite{25}. The distributions of wind forecast errors could be further refined by conditioning on current power, time of day, season, etc, however this study used a fixed t-distribution that best matched errors aggregated over the full month of July.

The actual loads and wind production data used in our experiments are shown in Figure \ref{fig:loads}. There are two important aspects of the data: first, the peak renewable generation is just over one quarter of the peak demand, corresponding to a moderate renewable energy penetration scenario. Second, there is a wind ramp event that occurs around midnight July 25th. The wind ramp event will pose an economic and reliability challenge to test the different sampling strategies in our experiments.

We examined the performance of each sampling strategy with 5, 10, 20, and 50 scenarios drawn at each 5 minute dispatch decision interval. The resulting costs of the first and second stage decisions for the full week are shown in Figure \ref{fig:costs} for all three strategies and the four sets of scenarios. Additionally, a summary of the costs for the full week is shown in Table \ref{tab:costs}. To make the differences in sampling strategies more clear, our model does not explicitly implement reserves. Instead any shortfall in generation results in an expensive loss-of-load penalty incurred during the second stage. This results in the bulk of the costs attributed to first stage dispatch decisions, and rare, but comparatively large second stage costs. In practice, these loss-of-load events would largely be addressed by ISO policies on reserve requirements.

Representative samples produced by each sampling technique, along with the nominal distribution, importance distri-
These scenarios allow the stochastic optimization algorithm to forecast values for wind from the importance distribution. For 20 or more scenarios, the Bayesian quadrature approach results in a consistent decrease in maximum second stage costs as the number of scenarios is increased. This consistent performance improvement is a strength of a quadrature-based approach. In contrast, importance sampling does not show a monotonic decrease in maximum second stage costs as more scenarios are considered. This is likely due to errors arising from approximating the true importance distribution.

With only 5 or 10 scenarios to consider, the importance sampling approach yields the lowest total costs. We attribute this performance to the additional information available from the deterministic solution. Recall that an estimate of the loss function $L(x, \xi)$ is required to determine the importance distribution. We approximate the loss function by solving a deterministic dispatch problem by assuming a perfect wind forecast. We then use the trapezoidal rule to calculate the normalizing factor that makes the importance distribution a proper pdf. Although this is a flawed estimate of the loss function, it is still valuable when only a few scenarios are available for optimization.

For 20 or more scenarios, the Bayesian quadrature approach yields a lower total cost. At this level of sampling, the Gaussian process approximation of the loss function is more accurate than the deterministic approximation made in importance sampling. We also note that the Bayesian quadrature approach results in a consistent decrease in maximum second stage costs as the number of scenarios is increased. This consistent performance improvement is a strength of a quadrature-based approach. In contrast, importance sampling does not show a monotonic decrease in maximum second stage costs as more scenarios are considered. This is likely due to errors arising from approximating the true importance distribution.
Fig. 3. Costs of the economic dispatch decisions from the three sampling strategies for 5, 10, 20, and 50 scenarios. While the Bayesian quadrature approach triggers expensive second stage costs more often than importance sampling, it achieves lower total costs for 20 and 50 scenarios with much lower computational costs.

V. CONCLUSION

This paper introduced two new techniques, importance sampling and Bayesian quadrature, to create scenarios of wind forecast error events for use in solving stochastic economic dispatch via a two-stage recourse problem. The proposed techniques were compared against a Monte Carlo sampling strategy in a weeklong study of 5 minute economic dispatch using realistic generation, load, and wind forecast data from the RTS-GMLC test case and NREL’s WIND Toolkit dataset. Both importance sampling and Bayesian quadrature achieve a reduction in estimator variance as compared to the baseline Monte Carlo sampling, which ultimately yield lower total costs for the two-stage economic dispatch decisions. Furthermore, the scenarios selected by the improved strategies are drawn further out into the tails of the distribution of possible wind ramp events than with Monte Carlo sampling. The resulting dispatch decisions improve grid reliability and robustness by finding feasible solutions that reduce loss of load, while still maintaining economically optimal dispatch decisions.

We also found that the computational cost of determining sample locations is lower for Bayesian quadrature. Impor-
tance sampling suffers from the curse of circularity in that the normalization factor to make the importance distribution a proper probability density function requires knowing the expectation one is trying to estimate in the first place. To overcome this, we solve a deterministic economic dispatch problem to estimate the loss function and then numerically integrate it to approximate the normalization constant. This additional solve increases the computation cost of importance sampling relative to Bayesian quadrature. Furthermore, in the Bayesian quadrature approach the sample points can be computed once and reused at successive timesteps because it is a true quadrature approach rather than a stochastic sampling approach. The Bayesian quadrature samples only depend on the nominal wind pdf and the Gaussian process kernel, allowing them to be computed offline for wind distributions that are specific to a time of day, season, or current power output. Our analysis simply used a Student’s t-distribution to represent the distribution of wind ramps over the course of a year, however more condition-specific distributions would likely improve performance even further.

The most economic dispatch decisions are found when using importance sampling with ten or fewer scenarios, and with Bayesian quadrature for more than ten scenarios. We attribute this to the extra information about the loss function that is gained through its deterministic approximation while computing the importance distribution. With more than ten scenarios, the Bayesian quadrature approach can develop a better loss function surrogate and find a lower cost dispatch strategy. While both proposed strategies improve the reliability from the baseline, the most economic strategy is not necessarily the most reliable. The solutions found via Bayesian quadrature for 20 and 50 scenarios produce the lowest total cost as shown in Table I, but as shown in Table III, they involve more recourse or second stage costs than the less economic strategies found via importance sampling, reflecting increased number of expected loss of load events.

Applying these strategies to a realistic test case of 5 minute economic dispatch with time varying loads, stochastic wind generation, and realistic costs yields improvements in total cost and in reliability represented by the second stage recourse costs. Depending on the number of scenarios used, total costs are reduced by 4.8% - 0.6% relative to the Monte Carlo baseline. A large component of the improvements in overall economic optimality is actually due to improvements in second stage costs that represent consumption of reserves or loss of load. The economically optimal dispatch strategy reduces these costs by 79.6% - 12.6%, demonstrating significant benefits to grid reliability and reserve requirements. Employing these advanced scenario generation strategies can significantly reduce costs and improve the reliability of operating the grid with high levels of stochastic renewable energy production.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08GO28308 with Alliance for Sustainable Energy, LLC, the Manager and Operator of the National Renewable Energy Laboratory. Funding provided by the Department of Energy’s Grid Modernization Lab Consortium and Exascale Computing Project.

REFERENCES

[1] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee, and M. Antieuscu, “A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation,” IEEE Transactions on Power Systems, vol. 26, no. 1, 2011.

[2] A. Soroudi and T. Anurace, “Decision making under uncertainty in energy systems: State of the art,” Renewable and Sustainable Energy Reviews, vol. 28, pp. 376–384, 2013.

[3] A. Shapiro, D. Dentcheva, and A. Ruszczyński, Lectures on stochastic programming: modeling and theory. SIAM, 2009.

[4] J. R. Birge and F. Louveaux, Introduction to stochastic programming. Springer Science & Business Media, 2015.

[5] A. Papavasiliou, S. S. Oren, and R. P. O’Neill, “Reserve requirements for wind power integration: A scenario-based stochastic programming framework,” IEEE Transactions on Power Systems, vol. 26, no. 4, 2011.

[6] J. Cheng, R. L.-Y. Chen, H. N. Najm, A. Pinar, C. Safa, and J.-P. Watson, “Chance-constrained economic dispatch with renewable energy and storage,” Computational Optimization and Applications, vol. 70, no. 2, pp. 479–502, 2018.

[7] M. Lubin, Y. Dvorkin, and S. Backhaus, “A robust approach to chance constrained optimal power flow with renewable generation,” IEEE Transactions on Power Systems, vol. 31, no. 5, pp. 3840–3849, 2016.

[8] C. Safa, R. L.-Y. Chen, H. N. Najm, A. Pinar, and J.-P. Watson, “Efficient uncertainty quantification in stochastic economic dispatch,” IEEE Transactions on Power Systems, vol. 32, no. 4, 2017.

[9] A. Staad, J.-P. Watson, R. J.-B. Wets, and D. L. Woodruff, “Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators,” Wind Energy, vol. 20, no. 12, 2017.

[10] A. J. Wood and B. F. Wollenberg, Power generation, operation, and control. John Wiley & Sons, 2012.

[11] J. Zha, Optimization of power system operation. John Wiley & Sons, 2015, vol. 47.

[12] J. A. Taylor, Convex optimization of power systems. Cambridge University Press, 2015.

[13] A. J. Conejo, M. Carrion, and J. M. Morales, Decision making under uncertainty in electricity markets. Springer, 2010, vol. 1.

[14] G. Giebel, L. Landberg, G. Kariniotakis, and R. Brownsword, “State-of-the-art methods and software tools for short-term prediction of wind energy production,” in EWEC 2003 (European Wind Energy Conference and exhibition), 2003.

[15] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, “Current methods and advances in forecasting of wind power generation.” Renewable Energy, vol. 72, no. 1, pp. 1–8, 2012.

[16] J. A. Bucklew, Introduction to Rare Event Simulation, ser. Springer Series in Statistics. New York, NY: Springer New York, 2004.

[17] C. E. Rasmussen and C. K. I. Williams, Gaussian processes for machine learning. Cambridge, MA: MIT Press, 2006.

[18] Z. Ghahramani and C. E. Rasmussen, “Baysian Monte Carlo,” in Advances in Neural Information Processing Systems 15, S. Becker, S. Thrun, and K. Obermayer, Eds. MIT Press, 2003, pp. 505–512.

[19] A. O’Hagan, “Bayes-Hermite quadrature,” in Bayes-Hermite quadrature error density estimators,” Wind Energy, vol. 20, no. 12, 2017.

[20] P. Hennig, M. A. Osborne, and M. Girolami, “Probabilistic numerics and uncertainty in computations,” Proc. R. Soc. A, vol. 471, no. 2179, p. 20150142, Jul. 2015.

[21] Grid Modernization Lab Consortium, “Reliability test system - Grid Modernization Lab Consortium,” 2017, [Online; accessed 11-December-2017]. [Online]. Available: https://github.com/GridMod/RTS-GMLC.

[22] P. Chairman, M. Bhavaraj, B. Biggerstaff et al., “IEEE reliability test system: a report prepared by the reliability test system task force of the application of probability methods subcommittee,” IEEE Trans Power Appar Syst, vol. 98, no. 6, pp. 2047–2054, 1979.

[23] C. Grigg, P. Wong, P. Albrecht, B. M. Hodge, and J. McCaa, “The wind integration national dataset (wind) toolkit,” Applied Energy, vol. 151, 2015.

[24] B.-M. Hodge and M. Milligan, “Wind power forecasting error distributions over multiple timescales,” in Power and Energy Society General Meeting, 2011 IEEE. IEEE, 2011, pp. 1–8.