Probabilistic Balancing of Grid with Renewables and Storage

Weixuan Gao, *Student Member, IEEE*
Dept. of Civil and Environmental Engineering
Stanford University
Stanford, CA 94305
gaow@stanford.edu

Dimitry Gorinevsky, *Fellow, IEEE*
Dept. of Electrical Engineering, Stanford University
and Mitek Analytics LLC
Palo Alto, CA 94306
gorin@stanford.edu, dimitry@mitekan.com

***Abstract*—This paper develops probabilistic analysis approach to ensure generating capacity is sufficient to balance load (demand) in power grid. In the future grid, most generation will come from variable sources, such as solar and wind. This means the grid can be only balanced probabilistically. Variability of renewable generation requires using storage, which must be a part of the analysis. Further, mandated reliability of the power systems is very high. This means the probabilistic analysis must carefully consider the combined impact of several random factors and include extreme events. The paper presents numerical probabilistic calculus methodology addressing all mentioned issues. It is illustrated for example using actual ISO data.**

***Index Terms*—Loss of Load, Renewables, Storage, Probabilistic Analysis, Data-driven Modeling, Quantile Regression**

I. Introduction

This paper develops tools for probabilistic analysis of grid reliability. The problem is to ensure generation capacity is sufficient for balancing electricity demand. Industry uses well-established analysis methods, e.g., see [1]. Yet, increase in variable generation from renewable sources creates new issues. States of California and Hawaii push for 100% renewable energy by 2045, Massachusetts, by 2050. France, Germany, Denmark, and other EU states have similar plans.

Existing methods for analysis of grid based on dispatchable generation are unsuitable for grid dominated by variable generation and with substantial energy storage. Probabilistic analysis of the latter is complicated by electricity demand, wind, insolation, and storage being strongly coupled variables.

The primary application considered herein is grid planning on a regional level. The problem is to analyze the adequacy of the generation (and storage) capacity to support the demand. The same probabilistic analysis tools can be applied to utility operational problems such as setting reserves and electricity market trading. The tools could be also used for distribution systems with a variable generation that might become autonomous.

A. State of the Art

The established method for evaluating grid reliability is to assume generator outages are independent random variables and use discrete convolution for probabilistic modeling, see [1], [7]. In existing practice, other random factors are usually included by Monte Carlo analysis with the sampling of loads and using capacity credit for Variable Energy Resources (VER). Monte Carlo analysis extensions for a system with storage are considered in [11], [12].

Closer to the approach in this paper, in [2], operational reserves are analyzed by convolving distributions of plant outage, conditional load, wind and solar forecast errors. These are considered to be independent random variables, which might be inaccurate.

The convolution method is attractive for grid reliability problems, but its practical use is limited by several factors. First, probabilistic modeling of time series requires many years worth of data. This has been recently addressed for load forecasting by using Quantile Regression (QR) modeling [3], [9], [10]. NERC-mandated 1-in-10 reliability of the power system (loss of load probability below 0.3%) requires complementing QR with careful modeling of the distribution tails, see [9], [10]. Second, peak load is known to be correlated with wind, e.g., see [6]. It is also correlated with solar generation. Finally, storage charge/discharge is highly correlated with VER generation and demand.

B. Contributions

This paper introduces probabilistic calculus that allows addressing the mentioned limitation. The contributions of this paper are as follows.

First, the paper demonstrates how data-driven conditional probability distribution models can be used to analyze the reliability of grid power balancing.

Second, the developed probabilistic power balancing analysis methods are extended to the grid with storage by adding energy balancing analysis.

Third, this paper demonstrates example scenarios that illustrate the use of proposed methods for planning trades between wind and storage.

C. Outline

The paper outline is as follows. Section II introduces operations of numerical probabilistic analysis. Section III presents data-driven modeling methods illustrated by ISO New England data examples. Section IV applies the developed tools to power balancing. Finally, Section V demonstrates an application of the power and energy balancing methods to grid with substantial storage and wind generation.

II. Probabilistic Calculus

This section introduces operations of numerical probabilistic analysis. The operations allow combining random variables that represent demand, outages, wind, and storage for evaluating grid reliability in a computationally scalable way.

A. Distribution Sampling

Consider random variable u with known probability distribution. Numerical probabilistic analysis requires sampling of the Cumulative Distribution Function (CDF) F_u as

$$F_u[k] = \mathbf{P}(u \leq k\Delta h), \quad p_u[k] = F_u[k] - F_u[k-1], \tag{1}$$

where k is an integer sample number and p_u is the Probability Mass Function (PMF) for the sampled CDF F_u. If sampling step Δh is small enough, such approximate model is accurate. The CDFs of the original distribution and sampled distribution match exactly at the sample points. The CDF F_u can be interpolated between the samples to give a continuous distribution.

As an example, consider Bernoulli distribution

$$\mathbf{P}(u = 0) = q, \quad \mathbf{P}(u = h) = 1 - q. \tag{2}$$

Applying (1) to (2) yields the sampled Bernoulli PMF

$$p_u[k = 0] = q, \;\; p_u\left[k = \lceil h/\Delta h \rceil\right] = 1 - q, \tag{3}$$

where $p_u[k] = 0$, for all other k and $\lceil x \rceil$ is the least integer that is greater than or equal x.

B. Independent Variables

Sampled models allow effective computation of distributions for variable sums. Consider two independent random variables u and v sampled on the same uniform grid. As well known, the PMF of the sum $u + v$ is discrete convolution

$$p_{u+v}[\,\cdot\,] = p_u[\,\cdot\,] * p_u[\,\cdot\,]. \tag{4}$$

Consider an example of computing outage distribution for a set of power generating units. The outages are described by independent Bernoulli distributions of the form (2) with different parameters for each unit. The distribution of the total outage power can be computed as a convolution of the individual sampled Bernoulli distributions of the form (3).

This example describes a well-known approach used for power systems reliability planning, e.g., see [1]. It is computationally efficient because it avoids combinatorial complexity present in exact summation of N different Bernoulli variables.

The accuracy of the sampled approximation can be checked by observing convergence as the sampling step is reduced. To understand why the approach provides accurate results, consider the error $\tilde{u}$ of approximating u with a sampled variation $\hat{u}$. For any other sampled variable v we have

$$\max F_{\tilde{u}+v}[\,\cdot\,] \leq \max F_{\tilde{u}}[\,\cdot\,]. \tag{5}$$

Inequality (5) holds because

$$\max\left(F_{\tilde{u}}[\,\cdot\,] * p_v[\,\cdot\,]\right) \leq \max F_{\tilde{u}}[\,\cdot\,] \cdot \sum_k p_v[k], \tag{6}$$

where the sum in (6) adds to 1. According to (5), approximation errors are attenuated as they propagate through addition of independent variables.

C. Conditional Distributions

Consider a sampled conditional random variable $u|Z$, where independent variables (regressors) Z are in a finite state space with states Z_j that have probabilities $\mathbf{P}(Z_j)$. For example described in Subsection III-A there are 4,032 such states. For each Z_j, there is a conditional variable and associated PMF

$$u_j = u|Z_j, \;\; u_j \sim p_{u_j}[\,\cdot\,]. \tag{7}$$

Evaluating grid reliability requires computing expectations, such as Loss Of Load Expectation (LOLE), that can be represented through certain conditional distributions as

$$\mathbf{E}[u \leq 0] = \sum_{j=1}^{m} \sum_{u_j \leq 0} p_{u_j}[k] \cdot \mathbf{P}(Z_j), \tag{8}$$

where index k is limited to the sampled values $u_j \leq 0$.

Now consider two conditionally independent random variables u and v. This means $u_j = u|Z_j$ and $v_j = v|Z_j$ are independent for any given j. Conditional independence neither implies nor follows from the independence of variables u and v. For a sum $w = u + v$, conditional distribution $w|Z$ can be computed by convolving conditional PMFs for u and v at all Z_j to obtain the PMFs for $w_j = w|Z_j$

$$p_{w_j}[\,\cdot\,] = p_{u_j}[\,\cdot\,] * p_{v_j}[\,\cdot\,], \; (j = 1, \ldots, M). \tag{9}$$

D. Linearly Dependent Variables

Consider a linear model where random variable w is used as predictor of variable u

$$u = v + \gamma \cdot w, \tag{10}$$

where γ is a constant and v is prediction residual independent of w. Though variables u and w in (10) are dependent, their sum can be represented as a sum of two independent variables

$$u + w = v + (1 + \gamma) \cdot w, \tag{11}$$

Distribution of (11) can be computed through the convolution

$$p_{u+w}[\,\cdot\,] = (1 + \gamma) \cdot p_v[\,\cdot\,] * p_w[\,\cdot\,], \tag{12}$$

Now consider a conditional distribution, where

$$u|Z = v|Z + \gamma(Z) \cdot w|Z. \tag{13}$$

In (13), Z and w can be viewed as independent predictors of u, while v is prediction residual. Using notation of (7)

$$(u + w)_j = v_j + (1 + \gamma(Z)) \cdot w_j, \tag{14}$$

where the PMF can be computed similar to (12) as

$$p_{(u+w)_j}[\,\cdot\,] = (1 + \gamma(Z_j)) \cdot p_{v_j}[\,\cdot\,] * p_{w_j}[\,\cdot\,], \tag{15}$$

E. Quantile Model

Distribution model might be available in quantile form

$$\mathbf{P}(u \le Q_i) = q_i, \quad (i = 1, \ldots, m), \tag{16}$$

where Q_j are *quantiles* and q_j are *quantile levels* that are related to 50/50, 90/10, etc levels used in the industry. Let $\{u_i\}_{i=1}^N$ be independent samples of the random variable u. Then q_j in (16) can be empirical quantiles $q_j = n_j/N$, where n_j is the number of points in the set such that $u_i \le Q_j$.

For NERC one-in-ten requirement quantiles of interest have just 2-3 hourly samples per year available for modeling. For modeling of distribution tails, when $q \ll 1$ or $1 - q \ll 1$, consistent approach is suggested by Extreme Value Theory (EVT), see [8]. Power system data often follow Exponential or Generalized Pareto distributions predicted by EVT. Estimating 2-3 parameters of these distributions from the data allows extrapolating the tail into the quantiles where data are scarce.

F. Quantile Model Sampling

Consider computing sampled model $\{y_k, f_k\}_{k=1}^K$ of the form (1) from data (16). As first step, quantiles Q_j are interpolated on a uniform grid of values $u_k = k\Delta h$ to obtain quantile levels y_k. Direct interpolation does not work very well because differentiating the CDF to get PMF amplifies numerical errors. The approach that was found to work well is based on solving optimal smoothed approximation problem

$$\text{minimize}_F \ \|Y - F\|^2 + \rho \|D^3 F\|^2, \tag{17}$$

where $Y = \text{col}\{y_k\}_{k=1}^K$ is the resampled quantile data vector, F is vector of smoothed CDF values f_k, D is the first difference matrix, and ρ is a regularization parameter. The sought PMF values are $p_u[k] = f_k - f_{k-1}$. Smoothing filter (17) for CDF is related to the well-known Hodrick-Prescott filter, with additional differentiation in (17) to get the PMF from the CDF. Matrix D^3 is sparse tri-diagonal, therefore solution to (17) can be efficiently computed for large problem size K.

III. Data-Driven Models

A. Modeling Data Example

This section uses ISO New England (ISO-NE) service area data for 2016. These publicly available time series data include: load [4] and wind generation [5]. We model wind intensity as a non-dimensional variable with PDF supported on $[0, 1]$ interval. The wind generation in the examples is wind intensity (random variable) times wind nameplate capacity, which depends on the scenario. The distribution of wind intensity is estimated from historical ISO-NE data.

Modeled load (electricity demand) is conditional on vector $Z \in \Re^{45}$ that includes 12 binary regressors selecting calendar months, 7 regressors selecting weekdays, 24 regressors selecting hours of the day, and 2 selecting holiday or not. There are total of $12 \times 7 \times 24 \times 2 = 4,032$ states for Z. Conditional distribution for wind uses Months and Hours regressors only; wind does not depend on Weekday and Holiday.

B. Quantile Regression

Quantile model of distribution for random variable u conditional on regressors Z has the form

$$\mathbf{P}(u \ge \beta Z + \alpha) = q \tag{18}$$

Parameter vector β and scalar α in QR model (18) can be estimated by solving an LP problem, see [10] and references there. QR is linear in Z and has one parameter more than the number of regressors. For wind, regressor vector Z includes Month and Hour indicators, the total of $36 = 12 + 24$. This is 8 times fewer regression parameters and better statistical averaging than simple binning of the data for each month at each hour, which requires $288 = 12 \times 24$ bins. The difference is in QR assuming that month and hour impacts add up linearly.

C. Multiple QR

Consider Multiple Quantile Regression (Multiple QR) problem for building QR models from data set

$$\mathcal{D} \equiv \{u_i, Z_i\}_{i=1}^N \tag{19}$$

Consider matrices

$$U = [u_1, \ldots, u_N]^T \in \Re^{N,1} \tag{20}$$

$$Z = [Z_1, \ldots, Z_N]^T \in \Re^{N,m} \tag{21}$$

As discussed in [10], multiple QR can be formulated as sparse convex optimization problem in Second Order Cone Program (SOCP) form. Model parameters, scalars α_i and vectors β_i, correspond to QR quantile levels q_i $(i = 1, \ldots, m)$. They can be computed as minimizers for the following SOCP

$$\{\alpha_i, \beta_i\}_{i=1}^m = \arg\min \sum_{i=1}^m h(U - Z\beta_i - \alpha_i; q_i) \tag{22}$$

$$+ \lambda \sum_{i=2}^m \|\beta_i - \beta_{i-1}\|^2 + \mu \sum_{j=2}^{m-1} (\alpha_{j+1} - 2\alpha_j + \alpha_{j-1})^2,$$

$$h(Y; q) = \tfrac{1}{2}\|Y\|_1 + (q - \tfrac{1}{2})\mathbf{1}_N^T Y, \tag{23}$$

where quantiles q_i are on uniformly spaced grid on $(0, 1)$ interval and $\mathbf{1}_N$ is a vector column of N ones. The selection of regularization parameters λ and μ is discussed in [10].

D. Conditional Distribution Example

As an example, ISO-NE wind generation data described in Subsection III-A were used. Regressor vector Z included 24 Hours indicators and 12 Months indicators. As discussed in Subsection III-C, multiple QR model of wind on a uniform grid of q_i with pitch of 0.05 was estimated by solving SOCP problem (22) using one year worth of hourly data.

For each state Z_j, the multiple QR model for wind provides conditional sampled distribution in quantile form, see Subsection II-E. For each state Z_j, a sampled PMF was obtained by solving regularized resampling problem (17). Example PMF of wind conditional on 12 noon and July is shown in Figure 1

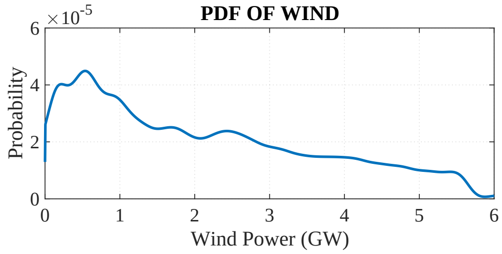

Fig. 1. Example PDF of wind intensity for 12 noon in July

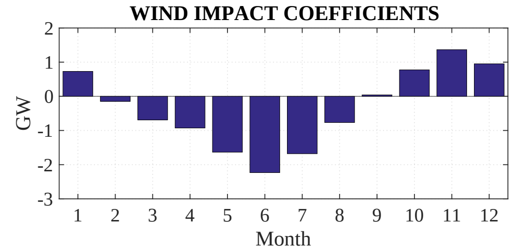

Fig. 2. Wind impact coefficients vs calendar month

E. Linear Dependence

Consider an extension of data set (19) with additional explanatory variable w

$$\mathcal{D}_W \equiv \{u_i, w_i, Z_i\}_{i=1}^N \tag{24}$$

In addition to (20)–(21), introduce matrix

$$V \quad = \quad [w_1 Z_1, \ldots, w_N Z_N]^T \in \Re^{N,m} \tag{25}$$

Multiple QR model for conditional distribution (13) can be estimated by solving the following extension of SOCP (22)

$$\{\{\alpha_i, \beta_i\}_{i=1}^m, \gamma\} = \arg\min \sum_{i=1}^m h(U - Z\beta_i - \alpha_i - V\gamma; q_i)$$
$$+\lambda \sum_{i=2}^m \|\beta_i - \beta_{i-1}\|^2 + \mu \sum_{j=2}^{m-1} \|\alpha_{j+1} - 2\alpha_j + \alpha_{j-1}\|^2$$
$$+ \nu \cdot \|D\gamma\|^2, \quad \text{subject to } C\gamma = 0 \tag{26}$$

Formulation (26) was used to estimate QR model for load dependence on wind from ISO-NE data for load and wind intensity described in Subsection III-A.

Linear conditional dependence of load on wind of the form (14) is assumed, where u is the load, and w is the wind intensity. The 45-indicator vector Z in the model is as described in Subsection III-A. Dataset (24) includes samples for u, w, and Z taken at every hour of one year.

Wind influences demand depending on the temperature. This is modeled by assuming wind influence factor γ depends on the calendar month. Components of γ corresponding to consecutive months are assumed to be close. To program this into (26), matrix $C \in \Re^{33,45}$ is a selector for components of γ corresponding to indicators of Hours, Weekdays, and Holidays, which leaves out Months. Matrix $D \in \Re^{12,45}$ in (26) provides circulant differences of γ components corresponding to Month indicators. Regularization parameter ν controls smoothness of the estimated dependence of γ on calendar month. Figure 2 shows wind impact coefficients γ depending on calendar month. Figure 3 shows load and model quantiles. At each point t, wind intensity $w(t)$ and time-dependent regressors $Z(t)$ are used to obtain load quantiles $Q_i = \beta_i Z(t) + \alpha_i + \gamma(Z(t)) \cdot w(t)$. Quantiles Q_i that correspond to quantile levels q_i in the plot legend are plotted along with the actual load.

Multiple QR model obtained from (26) covers quantile levels $q_1 = 0.025$ to $q_m = 0.975$ with pitch 0.05. For a

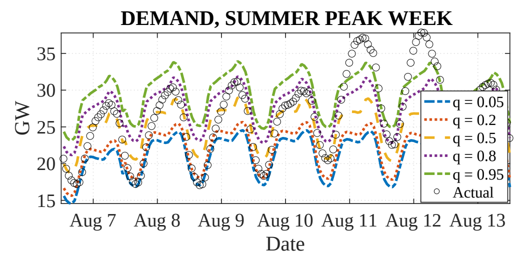

Fig. 3. System load and model quantiles for summer peak week

year worth of hourly data, there are about 200 points with $q \geq 0.975$ that are modeled as EVT tail, see Subsection II-E, by looking at Peaks Over Threshold (POT) data

$$e_{R,k} = y_{j_k} - Z_{j_k}\beta_m - \alpha_m, \tag{27}$$

where indexes j_k are such that $e_{R,k} > 0$. EVT Pareto tail model $q = \alpha(Q)$ is fitted to the POT data $e_{R,k}$, see [8], [9]. The sampling the tail model $\alpha(Q)$ yields an extension of the multiple QR model to the tail beyond quantile level q_m. For the tail, QR parameters β are assumed to be the same as β_m.

Similar to (27), the left tail is modeled based on POT data

$$e_{L,k} = y_{j_k} - Z_{j_k}\beta_1 - \alpha_1, \tag{28}$$

where indexes j_k are such that $e_{L,k} < 0$.

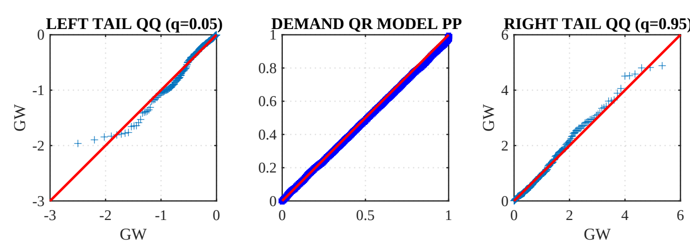

Fig. 4. QQ and PP plots for the load

Figure 4 illustrates model fit for the load data. The central panel shows PP plot of quantile level q in the QR model vs empirical quantile level defined in Subsection II-E. The left and right panels show QQ plots of the QR model quantile Q vs empirical quantile for the left and right tail respectively.

IV. Grid Balancing

This section applies the developed tools to power balancing assuming there is no storage. Storage is included in Section V.

A. Example System

As an example, consider power balancing for the ISO-NE service area. The example scenario is based on actual ISO-NE data for 2016 with nuclear capacity replaced by wind and increased fixed transfer capacity.

Load and wind data used for modeling are described in Section III-A. The model of wind intensity distribution is the same as described in Section III-D. The model is scaled by the assumed nameplate capacity of wind generation. Multiple QR load model is the same as illustrated in Figure 3.

B. Outages

Capacity and outage data for ISO-NE service area are reported into NERCs GADS database. The total $C_0 = 36.4$GW of dispatchable capacity is provided by 306 thermal generating units. Equivalent Forced Outage Rate - demand (EFORd) data give outage probabilities.

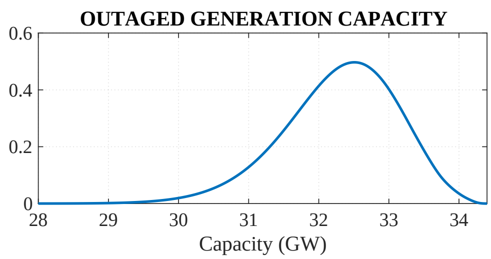

Fig. 5. Outage probability distribution

The GADS data provide unit capacity levels h and outage probabilities q in model (2). Sampled Bernoulli distribution models (3) were obtained from h, q and sampling step $\Delta h =$ 1MW. Outage PMF is a convolution of the sampled Bernoulli distributions for the individual outages, see Subsection II-B and [1]. Outage PMF is shown in Figure 5.

C. Reserve Margin

Reserve margin R is a random variable

$$R = F + C - O - L + W + S \tag{29}$$

where the r.h.s. terms are as follows:

- F is the fixed transfer capacity of 5GW,
- C is the dispatchable generation capacity of 36.4GW,
- O is the random variable giving total power capacity of the outage units with PMF shown in Figure 5,
- L is the load with conditional distribution model described in Subsection III-E,
- W is wind nameplate capacity $W_0 = 13.94$GW times wind intensity, which is modeled as described in Section III-D,
- S is used to model discharge power of energy storage in Section V and assumed zero in this section.

Consider distribution of R in (29) conditional on 45-indicator vector Z described in Subsection III-A. In (29), F, C, and S are constants. Random variable O is independent of L and W. Load L is linearly conditionally dependent on Wind W, similar to (13). Conditional distribution PMFs for $(-L) + W|Z$ can be computed in accordance with (15). To compute the conditional PMF of $R|Z$, negative outage $(-O)$ is then added as conditionally independent random variable see Section II-B and (9).

Described calculations consider distributions for outage O and wind W conditional on Z defined in Subsection III-A. For O, the PMF is the same for any Z. For W, the PMF is the same no matter what Holiday and Weekday indicators are.

The obtained sampled conditional probability distribution of reserve margin R (29) is illustrated in Figure 6 for Z corresponding to 12pm, Wednesday, July, Workday.

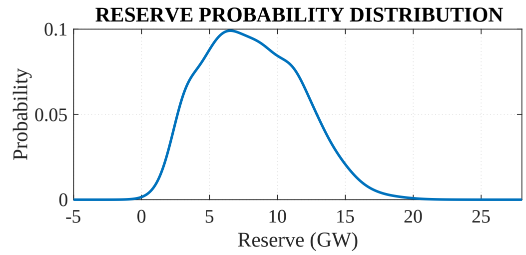

Fig. 6. PDF of reserve margin R for 12pm, Wednesday, July, Workday

D. LOLH Reliability Analysis

Negative reserve margin (29) R indicates that the grid cannot be balanced; there is not enough generating capacity. There are several ways of quantifying NERC one-in-ten requirement for capacity adequacy. This paper considers Loss of Load Hours (LOLH), the expected number of hours for the loss of load per year, and requires that LOLH ≤ 2.4.

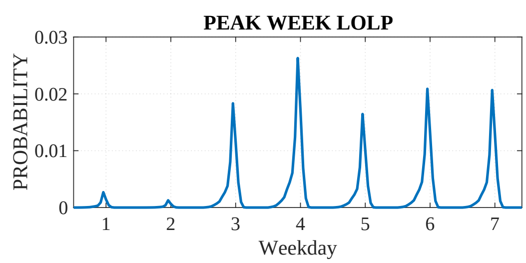

Fig. 7. Loss of Load Probability, LOLP$_j$, in the peak week

Conditional distribution of reserve margin allows computing Loss of Load Probability (LOLP), see Figure 7,

$$\text{LOLP}_j = \mathbf{P}(R \leq 0 \mid Z_j) = \sum_{R_j \leq 0} p_{R_j}[k]. \tag{30}$$

Computing LOLH in accordance with (8) and (30) yields

$$\text{LOLH} = 365 \times 24 \times \sum_{i=1}^{N} \text{LOLP}(Z_j) \cdot \mathbf{P}(Z_j) \tag{31}$$

States Z_j with Holiday indicated have probability $\mathbf{P}(Z_j) = p_H/2016$; the rest, $\mathbf{P}(Z_j) = (1 - p_H)/2016$, where $p_H = 9/365$ is the holiday probability.

V. Grid with Renewables and Storage

This section provides an extension of Section IV analysis, where energy storage is included.

A. Example Scenarios

Consider three scenarios based on ISO-NE data with scalings and modifications described below and in Table I. Scenario 0 with no energy storage was considered in Section IV. Scenario 1 has some storage and more wind generation, in Scenario 2 there is more storage. All scenarios yield LOLH of about 2 hours computed as described below.

Variable	*Scenario 0*	*Scenario 1*	*Scenario 2*
Dispatchable Capacity	34.4GW	34.4GW	34.4GW
Wind Nameplate Capacity	8.7GW	17.4GW	3.49GW
Energy Storage (4-hour)	-	2.5GW	3.5GW
Fixed Transfer	5GW	2GW	2GW
Load	37.8GW	37.8GW	37.8GW

TABLE I
Grid Planning Scenarios

Scenarios 1 and 2 assume storage with capacities $S_C = 2.5\text{GW}$ and $S_C = 3.5\text{GW}$ respectively that last $n_S = 4$ hours at peak generation. Probability distributions of wind intensity W, outage O, and load L are as described in Section IV.

B. Grid Planning

Power balancing problem for each scenario can be formulated as discussed in Section IV. This section adds a constant storage generation capacity assuming it is always available.

The last assumption holds if energy storage is never fully depleted. The new part of the analysis in this section is checking if the stored energy is depleted for each period of time. It is assumed that for 12 hours of the daily cycle, there is opportunity to recharge the storage to full capacity. This is a realistic assumption. It is then sufficient to check that the stored energy is not depleted for all periods of consecutive $1, 2, \ldots, 12$ hours without recharging.

C. Energy Margin

Energy reserve margin R_n for a period of consecutive n hours is the sum of reserve margins R (29) through these hours. It can be computed as convolution with rectangular window, $R_n = r_n * R$, where

$$r_n = [1, 1, ..., 1] \quad (32)$$

Applying the convolution to each term in (29) yields

$$R_n = F_n + C_n - O_n - L_n + W_n + S_n, \quad (33)$$

where the r.h.s. terms are as follows

- F_n is the fixed transfer capacity for energy, $F_n = n \cdot F$,
- C_n is the dispatchable energy capacity, $C_n = n \cdot C$,
- O_n is the energy outage; assuming outages for each our are independent, the PMF of O_n is n-fold convolution of outage PMF in Figure 5 with itself,
- L_n is energy demand variable, with conditional distribution estimated as described in Section III from the data $r_n * L$,
- W_n is wind generation variable, with conditional distribution estimated as described in Section III from the data $r_n * W$,
- S_n is the energy provided by the storage; assuming there was no charging, $S_n \leq \min(S_C, S_C \cdot n/n_S)$. Energy balance risk analysis assumes that $S_n = \min(S_C, S_C \cdot n/n_S)$.

For a given window width n in (32), calculation of conditional distribution $R_n|Z$ follows the methodology of Subsection IV-C to combine distributions of the r.h.s. terms in (33).

D. Planning Results

For conditional distribution $R_n|Z$, risk $\mathbf{P}(R_n \leq 0)$ and associated LOLH_n can be calculated as described in Subsection IV-D. Figure 8 shows computed risk of energy balancing LOLH_n vs n for Scenarios 1 and 2 in Table I. Note that LOLH_1 is the same as LOLH for power balancing.

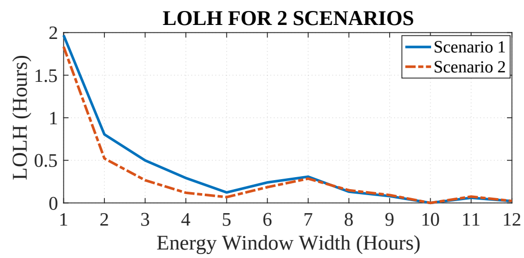

Fig. 8. Energy LOLH_n for Scenarios 1 and 2 vs. window width n.

For both scenarios LOLH_n is smaller for $n > 1$. This means power balancing risk dominates energy balancing risk. Scenario 2 shows about the same LOLH as Scenario 1: adding 1GW of storage replaces 13.9GW of wind nameplate capacity.

References

[1] Allan, R.N. and Billington, R., *Reliability Evaluation of Power Systems*, Springer Science & Business Media, 2013.

[2] Bucksteeg, M., Niesen, L., and Weber, C., "Impacts of dynamic probabilistic reserve sizing techniques on reserve requirements and system costs," *IEEE Trans. on Sust. Energy*, Vol. 7, No. 4, 2016, pp. 1408–1420.

[3] Hong, T., Wilson, J., and Xie, J., "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Trans. on Smart Grid*, Vol. 5, No. 1, 2014, pp. 456–462

[4] ISO-NE, *2016 SMD Hourly Data*. Available: https://www.iso-ne.com/isoexpress/web/reports/load-and-demand/-/tree/zone-info

[5] ISO-NE, *Hourly Wind Generation 2016*. Available: https://www.iso-ne.com/isoexpress/web/reports/operations/-/tree/daily-gen-fuel-type

[6] Keane, A., Milligan, M., Dent, C.J., et al., "Capacity value of wind power," *IEEE Trans. on Power Syst.*, Vol. 26, No. 2, 2011, pp. 564–572.

[7] Leite da Silva, A. M., Blanco, F. A. F. P., and Coelho, J., "Discrete convolution in generating capacity reliability evaluation - LOLE calculations and uncertainty aspects," *IEEE Trans. on Power Syst.*, Vol. 3, No. 4, 1988, pp. 1616–1624.

[8] Shenoy, S. and Gorinevsky, D., "Risk adjusted forecasting of electric power load," *Americal Control Conf.*, June 2014

[9] Shenoy, S., and Gorinevsky, D., "Data-driven stochastic pricing and application to electricity market," *IEEE Journ. Selected Topics in Signal Processing*, Vol. 10, No. 6, 2016, pp. 1029–1039.

[10] Shenoy, S., Gorinevsky, D., and Boyd S., "Non-parametric regression modeling for stochastic optimization of power grid load forecast," *Americal Control Conf.*, June 2015

[11] Sioshansi, R., Madaeni, S.H., and Denholm, P., "A dynamic programming approach to estimate the capacity value of energy storage," *IEEE Trans. on Power Systems*, Vol. 29, No. 1, 2014, pp. 395–403

[12] Sulaeman, S., Tian, Y., Benidris, M., and Mitra, J., "Quantification of storage necessary to firm up wind generation," *IEEE Trans. on Industry Applications*, Vol. 53, No. 4, 2017, pp. 3228–3236