APPLICATION OF SIGNAL PROCESSING METHODS IN ENERGY AND WATER SUSTAINABILITY OPTIMIZATION

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL ENGINEERING

AUGUST 2017

By

Seyyed Abolhasan Fatemi

Dissertation Committee :

Anthony Kuh, Chairperson

June Zhang

Anders Host-Madsen

Matthias Fripp

Reza Ghorbani

.

To Maryam, my loving wife

Acknowledgements

I would like to express my sincere gratitude to the following individuals who have helped me accomplish my dissertation for the degree. At the forefront, I would like to express my deep gratitude and appreciation to my dissertation supervisor, Professor Anthony Kuh for his continuous support and guidance throughout the dissertation stages. He has given his knowledge and put in effort at all times for the benefit of this dissertation.

I would like to thank Professor Matthias Fripp for many helpful discussions. I wish to thank other members of my dissertation committee: Professor Anders Host-Madsen, Professor Aleksandar Kavcic, Professor June Zhang, and Professor Reza Ghorbani for generously offering their time, support, guidance and good will throughout the preparation and review of this document. I also want to thank Professor Vijay Gupta from University of Notre Dame and Professor Makena Coffman for sharing their knowledge and ideas.

I like to thanks other faculty and staff that have played a meaningful role in my education, and students who made my workplace fun and enjoyable.

I would also like to acknowledge the funding support I received for the following grants: NSF grant ECCS-098344, DOE grant DE-0E0000394, and the University of Hawaii REIS project.

I would also like to acknowledge the immense impact and support that my wife, Maryam, has given me; her encouragement, patience and understanding was beyond description.

Abstract

Solar irradiation is a non-stationary process with its mean and variance changing depending on time and day of the year. For solar irradiation prediction we remove both seasonal and time of day effects to make observations approximately stationary and then use prediction methods. We propose to use the zenith angle (the angle between sun beam and perpendicular line on horizontal surface) to remove both seasonal and time of day effects. Our simulations using least-squares (LS), time-varying least squares (TVLS), exponentially weighted recursive least squares (EWRLS) and one step estimation of second order statistics shows that using zenith angle normalization gives lower mean square error than traditional normalization by subtracting the mean and then dividing by deviation.

In electric power grids, generation must equal load at all times. Since wind and solar power are intermittent, system operators must predict renewable generation and allocate operating reserves to mitigate imbalances. If they overestimate the renewable generation during scheduling, insufficient generation will be available during operation, which can be very costly. However, if they underestimate the renewable generation, usually they will only face the cost of keeping some generation capacity online and idle. Therefore overestimation of renewable generation resources usually presents a more serious problem than underestimation. Many researchers train their solar radiation forecast algorithms using symmetric criteria like RMSE or MAE, and then a bias is applied to the forecast later to reflect the asymmetric cost faced by the system operator - a technique we call indirectly biased forecasting. We investigate solar radiation forecasts using asymmetric cost functions (convex piecewise linear (CPWL) and LinEx) and optimize directly in the forecast training stage. We use linear programming and a gradient descent algorithm to find a directly biased

solution and compare it with the best indirectly biased solution. We also modify the LMS algorithm according to the cost functions to create an online forecast method. Simulation results show substantial cost savings using these methods.

We also propose two parametric probabilistic forecast methods by using beta and two sided power distribution for predicting solar irradiation and evaluate their performance. To improve their performance metrics a combining procedure based on the beta transformed linear opinion pool is utilized. Our simulations show that these methods -despite the simple structure- can accurately describe the stochastic characterization of solar irradiation and effectively reduce its uncertainty. The proposed approach is robust and algorithms can be modified for other point forecast methods.

We also consider reliability of electric grid as a public good and we use an insurance policy to implement a benefit taxation mechanism that provides a framework to achieve optimal reliability levels. Finally, we examine energy efficient scheduling for pumping water in water supply networks. This is formulated as a nonconvex optimization problem and we find solutions and conduct simulations for small networks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Using oil and other fossil fuels result in emission of ash and CO2 to the atmosphere which results in global warming and environmental contamination. This has changed people's perception as new more efficient and more sustainable energy sources are being examined and more energy efficient technologies are being developed to reduce energy consumption. To this end the main core of this dissertation is also related to optimizing energy usage. In Chapter 2-4 we discuss solar forecasting methods which is a requirement for efficient integration of solar generation into the power grid. In Chapter 5 we discuss reliability of the electric grid using an insurance framework. Finally, Chapter 6 discusses an important problem for the energy / water nexus. We discuss energy efficient scheduling algorithms for pumping water in water networks.

In recent years photovoltaic (PV) generation has rapidly increased as it provides a competitive solution for sustainable electricity generation. For example the total PV generation capacity reached 40 GW in the United States by installing about 15 GWdc solar PV in 2016. The growth will continue and the total installed PV capacity is expected to triple

over next five years [1]. Solar generation is also growing globally and in terms of world solar capacity the 300 GW milestone was reached in 2016 and the total global installed PV capacity is expected to exceed 700 GW in 2021 [2].

The PV generation in the electric grid has limitations since the generated power is time-varying and intermittent. Prediction and control, large scale storage, and demand response are three means that can be used to resolve these limitations. Despite all the progress in the battery and other storage technologies, very large scale storage is not viable yet (economical or environmental limitations) [3].

Solar irradiation is a non-stationary process with its mean and variance changing depending on time of the day and day of the year. For example on a sunny day, the received power in horizontal surface is maximum at around noon and at the same time of the day expected solar radiation in summer is more than winter. For solar irradiation prediction we need to remove both seasonal and time of day effects. For this purpose using satellite data is proposed in [4] and [5] using statistics of historical records. In [6], the authors used Fourier series to remove the seasonal effects in global solar radiation.

In Chapter 2 we use two approaches for normalization, first we use zenith angle (angle between sun beam and perpendicular line on horizontal surface) at every time to remove both seasonal and time of day effects. In the second approach, we use 60 successive days to normalize this piece of data by subtracting mean followed by dividing by deviation at each time of the day. For each of the approaches we investigate several forecasting methods like least squares (LS), exponentially weighted recursive least squares (EWRLS) and second order statistics are used. The simulation result for different sites shows that normalization using zenith angle is a simple but effective way to remove seasonal and diurnal effects.

Traditionally, researchers who have studied the problem of forecasting solar radiation,

evaluate their methods using symmetric cost criteria like root mean square error (RMSE) or mean absolute error (MAE). However, In many applications the actual cost function is not symmetric for example grid operators have more concern about shortage of production rather than its abundance, i.e. overestimation of resources has more serious consequences than underestimation. So in the grid operator's view the cost function is not symmetric. Therefore, in Chapter 3 we discuss solar radiation forecasts using convex piecewise linear (CPWL) and linear-exponential (LinEx) as asymmetric cost functions which are better fitted to the grid operator problem. For each of these cost functions we used two scenarios, i.e. adding bias to an unbiased forecast or formulating a directly biased forecast which considers the cost function from the outset. Under CPWL cost the forecast is formulated as a linear program and for LinEx cost we formulated the problem as a convex optimization and solved it by a gradient descent algorithm.

A key to Chapter 3 is establishing a general framework where we can make fair comparisons between the performance of forecast methods using asymmetric cost functions and other approaches using symmetric cost functions for solar radiation forecasts. We show that a method which creates a biased forecast by optimizing directly against the asymmetric function gives a better result than adding an optimal ex post-bias to a forecast previously trained with a symmetric cost function. We compare both "directly biased" and "indirectly biased" forecasts to emphasize the importance of using a directly biased forecast in this problem.

For many decision makers point forecasts that only give a single value for prediction of a random variable is not enough and more information is required. Probabilistic forecasts can give a more complete stochastic characterization using probability distribution over future quantity (for example predicting a CDF function for a random variable). Consid-

ering the limitations of point forecasts, relatively few papers have proposed probabilistic forecast for solar radiation prediction. For example, [7] suggests a Bayesian method to estimate parameters of the predictive distribution which was assumed to be a modified gamma distribution. Three different stochastic differential equation models are proposed in [8]. These models are first fitted to a training data set and subsequently evaluated on a one-year test set. [9] uses a quantile regression forest to obtain a set of quantile values from observed solar power and corresponding numerical weather prediction. An analog ensemble method proposed in [10] which is based on a set of historical numerical weather forecasts and corresponding observed solar power. A combination autoregressive moving average (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models are used in [11] to obtain probabilistic forecast of solar irradiance from historical observations. Extreme learning machine (ELM) is used for regression purposes and are trained to obtain nonparametric probabilistic forecast by predicting quantiles [12]. Quantile regression is used in [13] as a competitive indirect method for solar forecast under asymmetric cost function. However, probabilistic solar forecast is still immature and requires more in depth research [14, 15].

Chapter 4 proposes two parametric probabilistic forecast methods using beta and standard two sided power distribution and improves their performance metrics by a combining procedure based on beta transformed linear opinion pool. The proposed approach can be used to extend the results of many different point forecast methods which are minimizing RMSE or MSE.

Significant penetration of intermittent generation like wind and solar in the power grid introduces concerns about reliability of the grid due to stochastic nature of these energy sources. On the one hand, these intermittent generation sources improve long term supply

security and help transition from depletable energy sources toward sustainable renewable sources, on the other hand, unpredicted fluctuations of intermittent generation may aggravate reliability of the system and even lead to blackouts. In chapter 5, we argue that reliability of electric grid is a public good and current pricing system where electricity and reliability service are bundled together and enforce costumers to pay for the bundled price is not efficient.

Bundled pricing creates a trade-off between immediate operating cost and the long term goal of reliability as electricity and reliability are bundled together. The Utility hardly receives any signal about consumers' preferences on reliability so the immediate cost of operation is the winner and reliability levels maintained (hopefully) at minimum mandated standards (for example N-1 secure constraint which requires the service is provided under any single contingency event). So the reliability upgrades are postponed to the future and the reliability of the system constantly decreases as infrastructure deteriorates and demand increases over time. The degradation of the system continues until a big blackout happens in the over stressed system and the deficiencies of the power system are highlighted requiring immediate substantial upgrades. The cycles of degradation and substantial upgrades create a double paradox effect since the societies who are used to high levels of reliability experience more losses due outages as they are less prepared for it and less concerned about it [16].

Therefore, in Chapter 5, we propose to decouple electricity price from the reliability. For collecting provision cost of reliability, we use an insurance framework to implement a benefit taxation mechanism that provides a framework to achieve optimal reliability levels.

Electricity cost of pumping is a large part of total operation costs of water supply networks [17]. In order to decrease this cost some research has been conducted in [18–22].

Different optimization algorithms (linear Programming [22, 23], quadratic programming, dynamic programming, genetic algorithm[24],ant colony[25] heuristics [26]), various hydraulic modeling (nonlinear hydraulic[27], simplified hydraulic[28, 29], mass transfer[30]), and different decision variables (pump operating time , tank levels) [31] have been explored. In most of the past research, users do not have any active participation and demands are either modeled as a deterministic model or a stochastic process that is predicted using a time series model.

In Chapter 6 we consider active participation for the users of a radial network and we show that suitable collaboration of users can reduce the required energy to pump the desired amount of water for all users. We use a nonlinear hydraulic model for the pump (both fixed speed and variable speed) and pipes and we find the most energy efficient schedule for two users using nonlinear optimization. The solution for the two user case, is used to find an approximate solution for more than two users.

The main contributions of this dissertation are using zenith angle for solar power radiation normalization, modeling asymmetric cost for renewable power forecast error and obtaining directly biased forecast for solar power, probabilistic forecast of solar power. Also an insurance framework for reliability of the power grid and methods for optimizing energy in water networks are given. These results are partially published in [13, 32–37].

Chapter 2

Forecasting Solar Irradiation Using Zenith Angle

2.1 Introduction

Solar irradiation is a non-stationary process with its mean and variance changing depending on time of the day and day of the year. For example on a sunny day, the received power in horizontal surface is maximum at around noon and at the same time of the day expected solar radiation in summer is more than winter. Solar irradiation prediction becomes simpler if we remove both seasonal and time of day effects. For this purpose using satellite data is proposed in [4] and [5] used statistics of historical records. In [6], the authors used Fourier series to remove the seasonal effects in global solar radiation.

In this chapter we use two approaches for normalization, first we use zenith angle (angle between sun beam and perpendicular line on horizontal surface) at every time to remove both seasonal and time of day effects. In the second approach, we use 60 successive days

to normalize this piece of data by removing mean and variance versus time of the day. For each of the approaches we investigate several forecasting methods like least squares (LS), exponentially weighted recursive least squares (EWRLS) and second order statistics. The simulation result for different sites shows that normalization using zenith angle is a simple but effective way to remove seasonal and diurnal effects. This chapter is organized as follows. In Section 2.2, forecast methods using zenith angle are explained. The same methods but using traditional normalization is discussed in Section 2.3. Section 2.4 presents the simulation results and discussion. A summary of results and conclusion is given in Section 2.5.

2.2 Forecasting using zenith angle

Solar irradiation is a random process which consists of some deterministic and some random parameters. One of the dominant deterministic parameters in solar radiation is sun position and in particular zenith angle. Zenith angle (θ_z) is the angle between sun beam and perpendicular line on horizontal surface. Every day the sun rises when $\theta_z = 90^\circ$ and gradually decreases until around noon and then starts to increase until again reaches $\theta_z = 90^\circ$ when the sun sets. Time of sunrise and sunset is not constant during a year, similarly, zenith angle at specific time of the day changes during a year. For example, zenith angle at winter noon time is larger than summer noon time. For more illustration, in Hawaii on Nov 20th the sun rises at 6:48 when $\theta_z = 90^\circ$ then decreases to 41.2° at 12:20 then increases until the sun sets at 17:48 when again $\theta_z = 90^\circ$. On the other hand, on May 26 (July 16) the sun rises at 5:53 (6:02) when $\theta_z = 90^\circ$ and decrease to 0° at 12:30 (12:39) then increases until the sun sets at 19:07 (19:16) when again $\theta_z = 90^\circ$.

Zenith angle is deterministic and calculated by position of earth around the sun. The formula for the zenith angle is

$$\cos\theta_z = \cos\phi\cos\delta\cos\omega + \sin\phi\sin\delta$$

where ϕ is latitude; ω is sun time which is negative in mornings, zero at noon and positive in afternoons, and changes by $15°$/hour rate; and $\delta = -23.45° \cos\frac{360(d+10)}{365}$ where d is day of year [38].

Each point in Fig. 2.1 represents one minute solar radiation versus cosine of solar zenith angle. The data is for Oahu in 2010 and is obtained from [39]. The figure shows high correlation between $\cos\theta_z$ and irradiation.

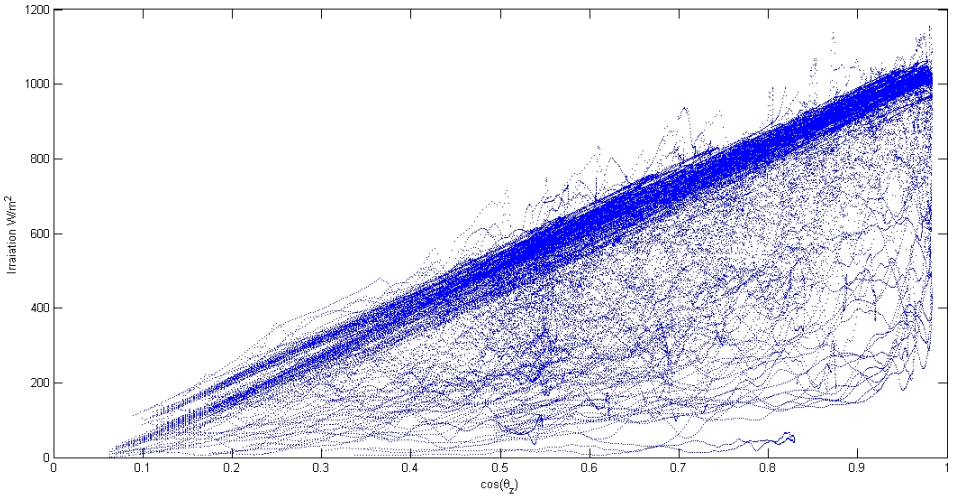

Figure 2.1: $\cos\theta_z$ and irradiation are highly correlated.

Since $\cos\theta_z$ is a measure of time we can use it instead of time index by a transform. This transform changes non-linear relationship between irradiation and time of the day to approximately linear relation between irradiation and $\cos\theta_z$. This transform also elimi-

nates seasonal effects caused by change of sun position.(Seasonal effects caused by clouds pattern will not be removed.) Fig. 2.2 shows a partly sunny day in summer, a sunny day in winter, and a day with sunny morning and cloudy afternoon as well as their transforms.

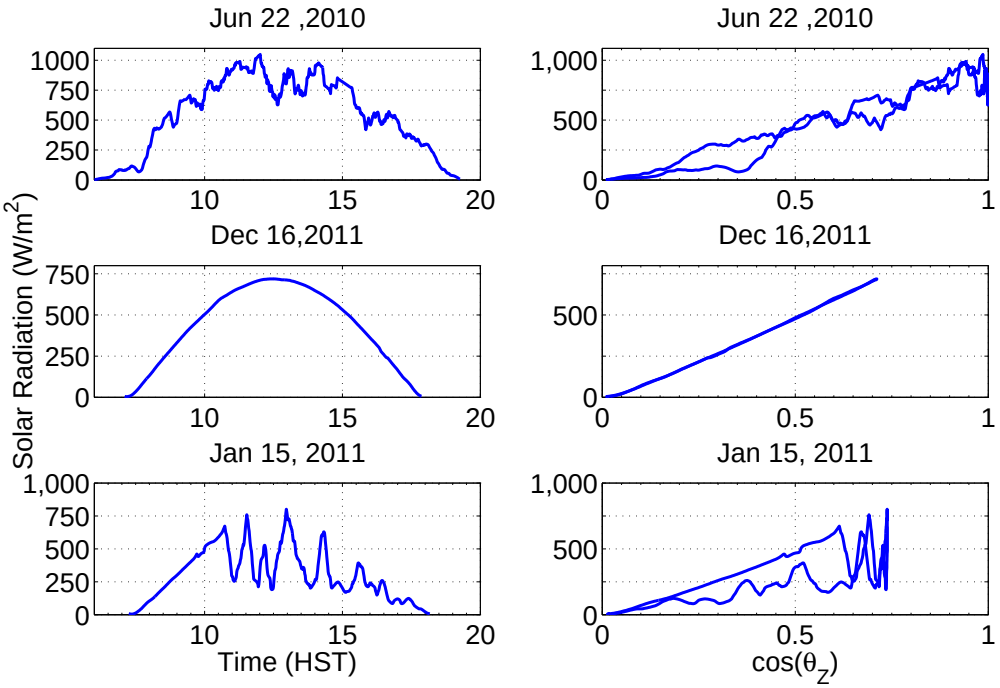

Figure 2.2: Transform of time to $\cos\theta_z$ (in sunny days relationship between irradiation and $\cos\theta_z$ is almost linear)

In this section we divide radiation at each time by corresponding $\cos\theta_z$ to decouple data from time. Let solar irradiation at time n be $r(n)$ and corresponding time decoupled data be $x(n)$

$$x(n) = \frac{r(n)}{\cos\theta_z(n)}$$

2.2.1 LS using zenith angle

Now we want to forecast $r(n+k)$ in which k varies from 10 to 120 minutes and sampling resolution is 1 sample per minute. We start Least Squares in which future forecast is estimated by linear combination of past data with specific constant weights. Here, we consider one of year data and divide it by corresponding $\cos\theta_z$. Then we find optimal weight vector W that predict future data using combination of past data. Let $y(n)$ be an $(m+1)\times 1$ vector of past decoupled data at time n converted to time of prediction by multiplication of $\cos\theta_z(n+k)$ and m is number of taps. Since $x(n)$ is not zero mean we need to add constant 1 to our vector

$$y(n) = [1, x(n), ..., x(n-m+2), x(n-m+1)]^T \times \cos\theta_z(n+k)$$

Let Y be matrix of training inputs, D be row vector of corresponding desired output, and N be number of data samples per day.

$$Y = [y(1), y(2), y(3), ..., y(365N)]$$
$$D = [r(1+k), r(2+k), r(3+k), ..., r(365N+k)]$$

If YY^T is of full rank optimum weight using least squares method is

$$W = ((YY^T)^{-1}Y)D^T$$

Finally predicted $r(n+k)$ is

$$\hat{r}(n+k) = W^T.y(n)$$

The computed weight vector for Oahu dataset is shown in following table **??**. As we use more taps the intercept decreases.

Table 2.1: Weight vectors

number of taps	W0	W1	W2	W3	W4	sum of weights without W0
single tap	394.8	0.487				0.487
10 taps	368.1	0.933	-0.537	-0.0008	-0.038	0.523
20 taps	345.4	0.858	-0.471	-0.0382	-0.078	0.554
40 taps	319.1	0.852	-0.445	-0.0265	-0.039	0.590
60 taps	305.1	0.831	-0.428	-0.0254	-0.037	0.609

Fig. 2.3 shows the resulted RMSE for an hour ahead forecast using LS and zenith angle in simulations for different sites. For each site we used one year for training and another year for testing to validate results. It reveals that increasing number of taps has little effect on improving performance.

In Fig. 2.4 the red solid line is measured irradiation in Oahu in the first week of April and the blue dashed line is an hour ahead forecasting using single tap LS and zenith angle. Fig. 2.5 shows that forecasting error rises as time of horizon of prediction increases.

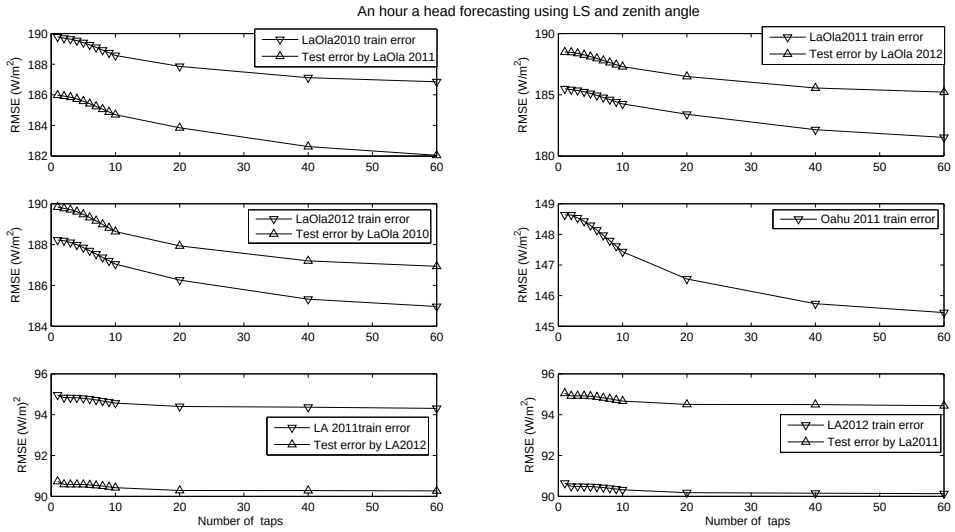

Figure 2.3: Increasing number of taps improves the performance some

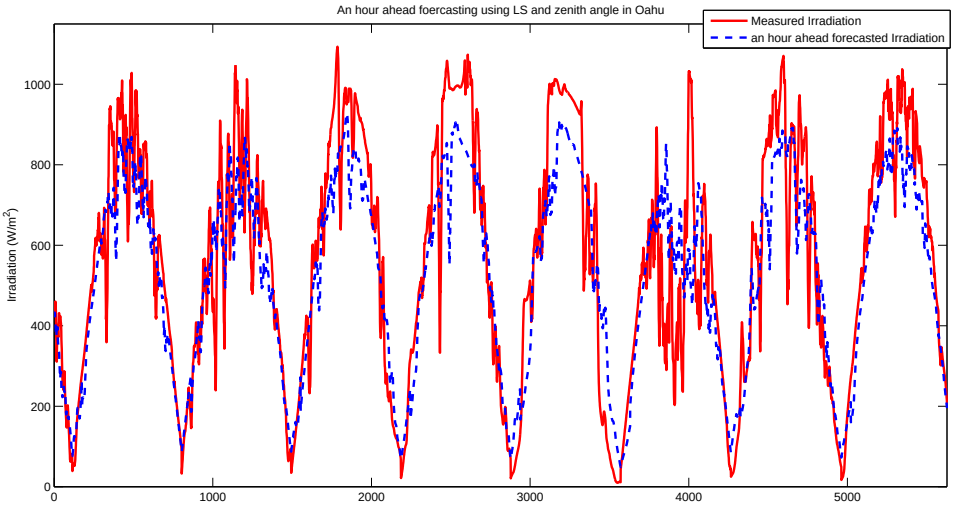

Figure 2.4: Sample forecast using LS and zenith angle in Oahu

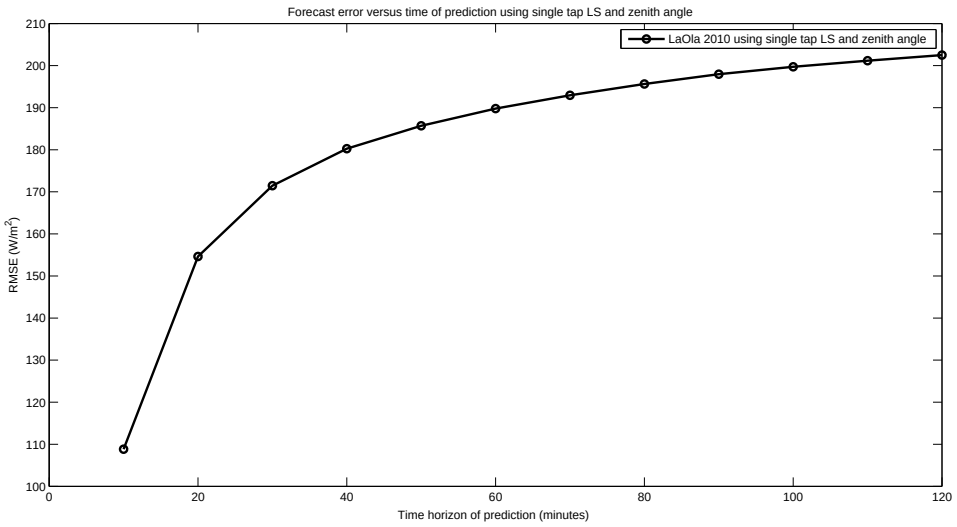

Figure 2.5: Forecasting error versus time horizon of prediction

2.2.2 Time varying least squares using zenith angle

Since using zenith angle does not remove seasonal effect related to clouds, it is reasonable to check if the results improve by using smaller training sets. Thus we check time varying Least Squares (TVLS). While in Least Squares the weights are constant during all times of year, in TVLS weights are changed at each time update. Training data in Least Squares is total data for one year, however, training data in TVLS for each time is a frame from a previous time up to present. Let L be length of training frame

$$Y(n) = [y(n-k-L+1), y(n-k-L+2), y(n-k-L+3), ..., y(n-k)]$$
$$D(n) = [r(n-L+1), r(n-L+2), r(n-L+3), ..., r(n)],$$

then the solution is given by

$$W(n) = ((Y(n)Y(n)^T)^{-1}Y(n))D(n)^T$$

and

$$\hat{r}(n+k) = W(n)^T.y(n)$$

As shown in Figure 2.6 single tap TVLS using zenith angle needs many training point to be as accurate as LS using zenith angle. We get similar results using more taps. Although in LS we use one year data for training however calculations are done only one time, but in TVLS computation is done at every time thus complexity of calculation dramatically grows as number of training points increases. In the next subsection we resolve these issues using EWRLS.

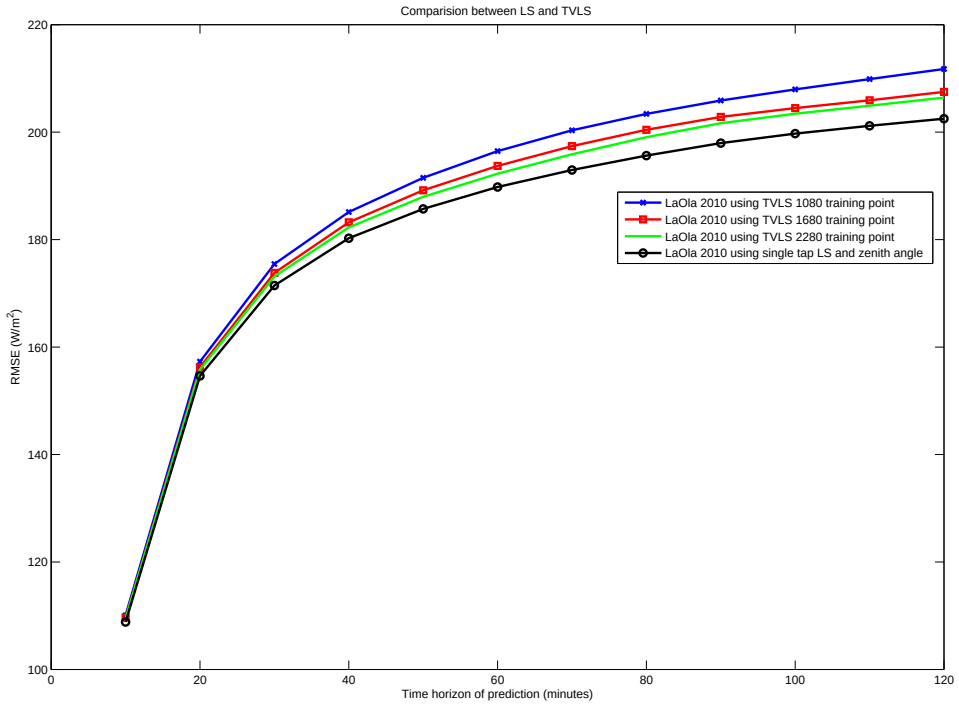

Figure 2.6: Increasing number of training points (L=1080-etc...) improves the result of TVLS using zenith angle however it needs many training points to be as precise as LS using zenith angle.

2.2.3 Exponentially Weighted Recursive Least Squares (EWRLS) by using zenith angle

As we found that larger training frame may give better result but need more complex computation, we tried to implement Exponentially Weighted Recursive Least Squares (EWRLS). RLS is a recursive algorithm for finding least squares solution [40]. The algorithm is able to solve exponentially weighted linear least squares by choosing an appropriate forgetting factor. While TVLS requires to calculate and invert $Y(n)Y(n)^T$ and multiply it by $Y(n)$ and $D(n)^T$, in EWRLS we just use new data to correct previous result. In this way, using EWRLS has both less computation cost and more accuracy comes from larger training frame.

Let

$$d(n) = r(n)$$

$$y(n) = [1, x(n), x(n-1), ...,x(n-m+2), x(n-m+1)]^T \times \cos\theta_z(n+k)$$

The weight vector ($W(n)$) updates at each step by following equations

$$W(n) = W(n-1) + K(n)(d(n) - y(n-k)^T W(n-1))$$

where $K(n)$ is $m \times 1$ gain vector

$$K(n) = \frac{\Sigma(n-1)y(n-k)}{\lambda + y(n-k)^T\Sigma(n-1)y(n-k)}$$

Σ is $m \times m$ covariance matrix

$$\Sigma(n) = \frac{1}{\lambda}(\Sigma(n-1) - K(n)y(n-k)^T\Sigma(n-1))$$

and λ is forgetting factor.

Finally

$$\hat{r}(n+k) = W(n)^T y(n)$$

Results of EWRLS using zenith angle is slightly better than LS using zenith angle. Figure 2.7 compares the results of single tap filter using EWRLS vs using LS. Similar results is obtained for more taps.

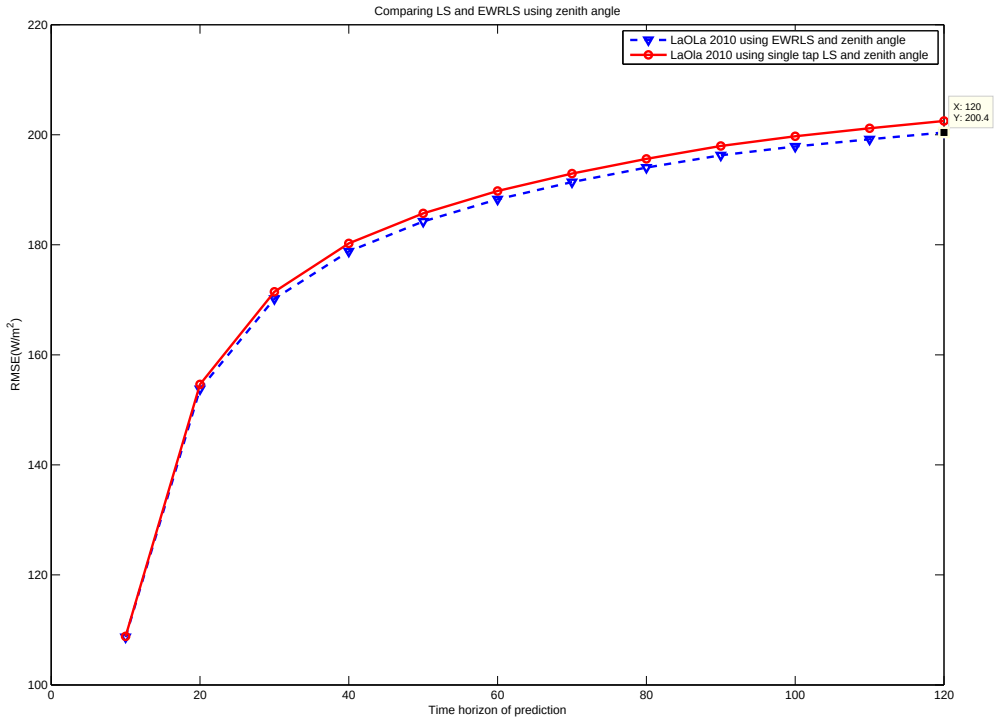

Figure 2.7: EWRLS using zenith angle is slightly better.

2.2.4 Forecasting using one step second order statistics

This method comes from linear mean square error estimator (LMMSE) which needs mean and covariance. Here, we use mean and auto-covariance. This method is easy to implement. Let λ be forgetting factor then mean and auto-covariance updated at each step by following equations:

$$\bar{x}(n) = \lambda\bar{x}(n-1) + (1-\lambda)x(n)$$

$$C_{xx}^{n}(0) = \lambda C_{xx}^{(n-1)}(0) + (1-\lambda)(x(n) - \bar{x}(n))^2$$

$$C_{xx}^{n}(k) = \lambda C_{xx}^{(n-1)}(k) + (1-\lambda)(x(n-k) - \bar{x}(n))(x(n) - \bar{x}(n))$$

Then

$$\hat{x}(n+k) = \bar{x}(n) + \frac{C_{xx}^{n}(k)}{C_{xx}^{n}(0)}(x(n) - \bar{x}(n))$$

Then we find the predicted output by considering θ_z at time of forecasting.

$$\hat{r}(n+k) = \hat{x}(n+k)\cos\theta_z(n+k)$$

Simulation of this method in LaOLa showed that this method could be as accurate as EWRLS using zenith angle with lower computation (Figure 2.8)

2.3 Normalizing data to be zero mean unit variance by removing time of day average and variance

In order to evaluate effectiveness of using zenith angle for forecasting, in this section we normalize data by removing mean and variance versus time of the day. Since seasonal effects happen gradually during a year instead of using several years data records

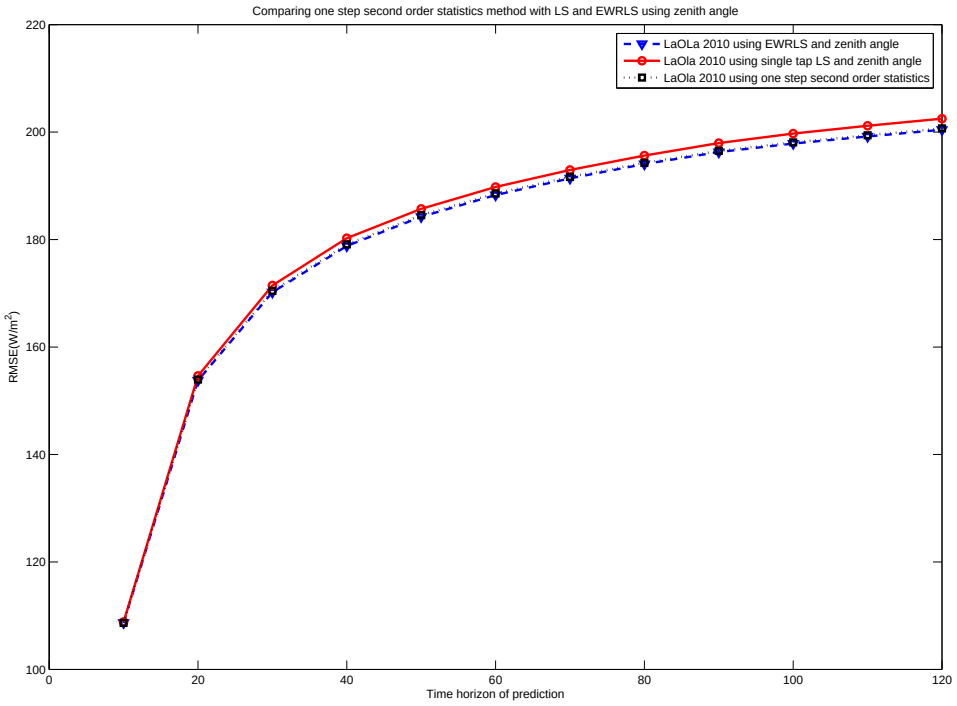

Figure 2.8: In simulation of LaOla using one step second order statistics using zenith angle is as accurate as EWRLS using zenith angle

for estimating mean and variance of irradiation at specific time of particular day, we can approximate that mean and variance using some neighbor days at that specified time of the day. There is a trade off in selecting number of neighbor days, if we consider more days we can estimate mean and variance more precisely, while we have less accuracy since we neglect seasonal effects more. Let N be number of samples per day and $r(n)$ be measured radiation at time n, we define mean and variance of radiation at same time of day in m_s past days i.e.

$$\mu_r(n) = \frac{1}{m_s + 1} \sum_{i=0}^{m_s} r(n - Ni)$$
$$\sigma_r^2(n) = (\frac{1}{m_s + 1} \sum_{i=0}^{m_s} r^2(n - Ni)) - \mu_r^2$$

Now our normalized data is $x(n)$

$$x(n) = (r(n) - \mu_r(n))/(\sigma_r(n))$$

Here we used $m_s = 60$.

2.3.1 Least Squares (LS)

Now we want to forecast $x(n + k)$ where k vary from 10 to 120 minutes and sampling resolution is 1 sample per minute. Our first attempt is training data using Least Square in which future forecast estimated by linear combination of past data with specific constant weights.

$$\hat{x}(n + k) = W^T y(n)$$

where W is an $m \times 1$ vector of weights and $y(n)$ is an $m \times 1$ vector of past data and m is number of taps.

$$y(n) = [x(n), x(n-1), ..., x(n-m+2), x(n-m+1)]^T$$

For finding optimal weights using LS we need training data. For this purpose we normalize one year data using above method. Let N be number of samples per day,Y be matrix of training inputs and D be row vector of corresponding desired output

$$Y = [y(1), y(2), y(3), ..., y(365N)]$$

$$D = [x(1+k), x(2+k), x(3+k), ..., x(365N+k)]$$

If YY^T is of full rank, optimum weight using least squares is

$$W = ((YY^T)^{-1}Y)D^T.$$

Finally

$$\hat{x}(n+k) = W^T y(n)\hat{r}(n+k) = \sigma_r(n+k-N)\hat{x}(n+k) + \mu_r(n+k-N).$$

Increasing number of taps marginally improves result. For example, for forecasting one hour ahead in all stations in Figure 2.9, comparing RMSE of 60-tap vs one-tap, we have less than 1.5% improvement.

2.3.2 Time Varying Least Squares (TVLS) Method

We have discussed TVLS using zenith angle in subsection 2.2.2. Here we use the same method but using conventional normalization. Let L be length of training frame

$$Y(n) = [y(n-k-L+1), y(n-k-L+2), y(n-k-L+3), ..., y(n-k)]$$

$$D(n) = [x(n-L+1), x(n-L+2), x(n-L+3), ..., x(n)]$$

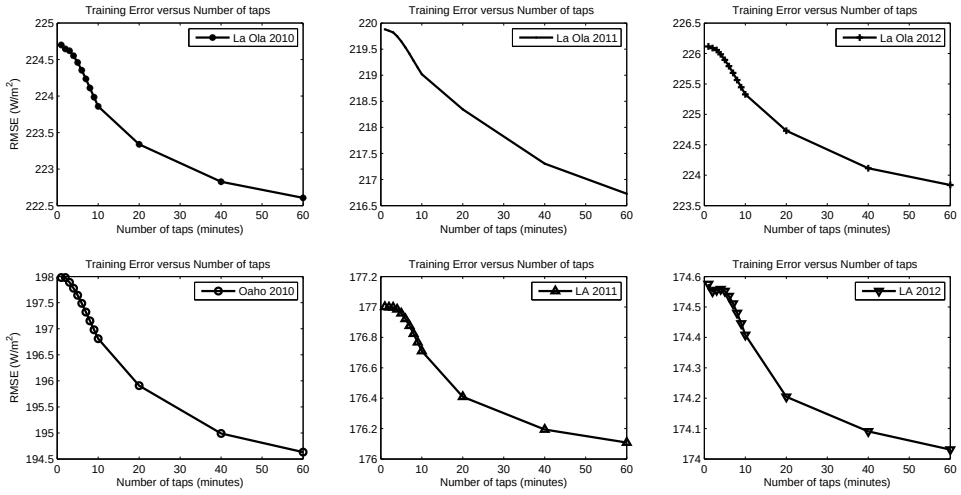

Figure 2.9: Increasing number of taps improves training error very slightly.(one hour ahead forecast)

which results in,

$$W(n) = ((Y(n)Y(n)^T)^{-1}Y(n))D(n)^T.$$

Then,

$$\hat{x}(n+k) = W(n)^T y(n)$$

$$\hat{r}(n+k) = \sigma_r(n+k-N)\hat{x}(n+k) + \mu_r(n+k-N).$$

Here, simulation results of TVLS is more precise than LS. Figure 2.10 shows although increasing number of taps form one to two give more accuracy for less than 30 minutes, more than two taps (even 10 taps) did not result in more significant improvement. Another effective parameter in this method is size of training frame. Increasing size of training data improves forecasting error ,however, it increases computation complexity. Figure 2.11 shows this fact via simulations of LaOLa 2010 data.

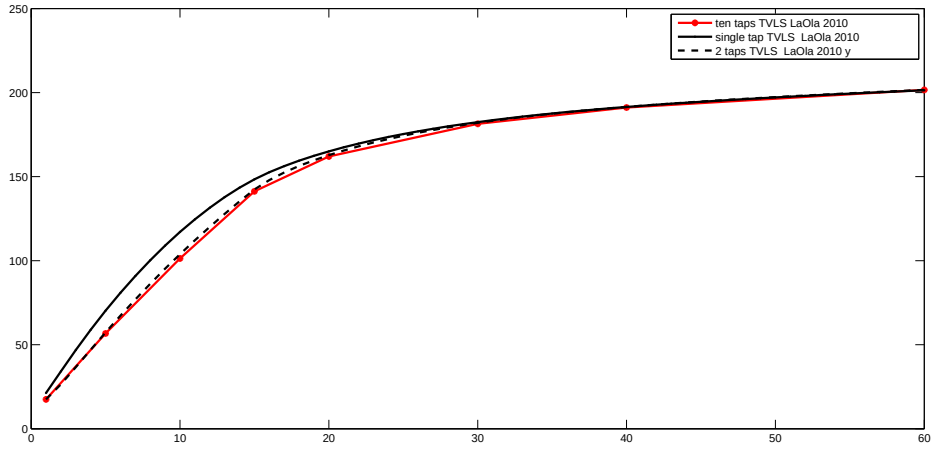

Figure 2.10: Increasing number of taps more than two does not have significant effect

2.3.3 Exponentially Weighted Recursive Least Squares (EWRLS)

As we found that larger training frame gives better results but need more complex computation, we tried to implement Exponentially Weighted Recursive Least Square(EWRLS). While TVLS requires to calculate and invert $Y(n)Y(n)^T$ and multiply it by $Y(n)$ and $D(n)^T$, in EWRLS we just use new data to correct previous result. In this way, using EWRLS has both less computation cost and more accuracy comes from training frame.

Let

$$d(n) = x(n)$$

$$y(n) = [x(n), x(n-1), ...,x(n-m+2), x(n-m+1)]^T$$

The weight vector ($W(n)$) updates at each step by following equations

$$W(n) = W(n-1) + K(n)(d(n) - y(n-k)^T W(n-1))$$

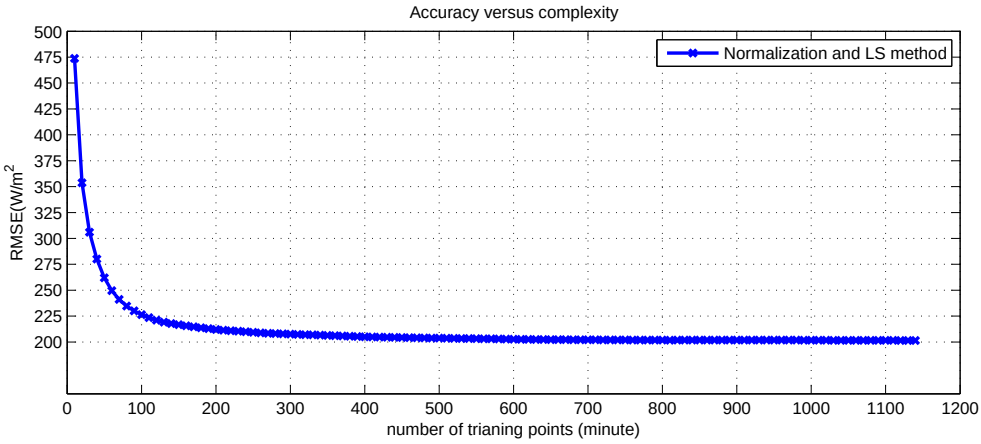

Figure 2.11: Increasing size of training frame improves accuracy of prediction but increases complexity.

where $K(n)$ is $m \times 1$ gain vector

$$K(n) = \frac{\Sigma(n-1)y(n-k)}{\lambda + y(n-k)^T \Sigma(n-1)y(n-k)}$$

Σ is $m \times m$ covariance matrix

$$\Sigma(n) = \frac{1}{\lambda}(\Sigma(n-1) - K(n)y(n-k)^T \Sigma(n-1))$$

and λ is forgetting factor. Here, we use $\lambda = 0.9997$ which gives the best result.

Finally

$$\hat{x}(n+k) = W(n)^T y(n)$$

$$\hat{r}(n+k) = \sigma_r(n+k-N)\hat{x}(n+k) + \mu_r(n+k-N)$$

As you can see in figures 2.12,2.13,and 2.14 using EWRLS slightly improves the result respect to LS using normalization. Comparison vs this method using zenith angle will be given in section 2.4.

2.3.4 Forecasting by one step estimation of second order statistics

Since we find from previous discussion that we can forecast only by single tap, here we predict by estimating mean and one step auto-covariance. This method is easy to implement:

$$\hat{x}(n+k) = \bar{x}(n) + \frac{C_{xx}^{n}(k)}{C_{xx}^{n}(0)}(x(n) - \bar{x}(n))$$

Where

$$\bar{x}(n) = \lambda\bar{x}(n-1) + (1-\lambda)x(n)$$

$$C_{xx}^{n}(0) = \lambda C_{xx}^{(n-1)}(0) + (1-\lambda)(x(n) - \bar{x}(n))^2$$

$$C_{xx}^{n}(k) = \lambda C_{xx}^{(n-1)}(k) + (1-\lambda)(x(n-k) - \bar{x}(n))(x(n) - \bar{x}(n))$$

Then denormalize data by using of yesterday average and variance at time of prediction.

$$\hat{r}(n+k) = \sigma_r(n+k-N)\hat{x}(n+k) + \mu_r(n+k-N)$$

2.4 Comparing methods

For evaluating methods we used the following datasets which are downloaded from www.nrel.gov/midc. These data have 1 minute resolution.

Hawaii Kalaeloa Oahu,	Latitude:21.31347 N Longitude:158.08257 W Elevation: 11 meters AMSL	4/1/2010-3/31/2011
Hawaii La Ola Lanai	Latitude:20.76685 N Longitude:156.92291 W Elevation: 381 meters AMSL	1/1/2010-12/31/2010 1/1/2011-12/31/2011 1/1/2012-11/31/2012
Los Angeles California	Latitude:33.966674 N Longitude:118.42282 W 27 meters AMSL	1/1/2011-12/31/2011 1/1/2012-11/31/2012

Simulation result of methods in these sites shows that by LS, EWRLS, and using one step second order statistics are more effective methods of forecasting using zenith angle normalization than using traditional normalization by removing mean and dividing by deviation.

Effectiveness of methods depends on weather characteristic of the site and desired time horizon of forecasting, for example, the RMSE in Los Angles data sets is lower than corresponding forecast for LaOla and Oahu. This is intuitively justified as Los Angeles has more sunny days than Hawaii this results in more easily predicted solar irradiation values for Los Angeles. In simulation of LaOla the best results were achieved using EWRLS and one step second order statistics with zenith angle. In Oahu station best result comes from LS with zenith and EWRLS with zenith angle. However, for LA station the best result comes from LS using zenith.

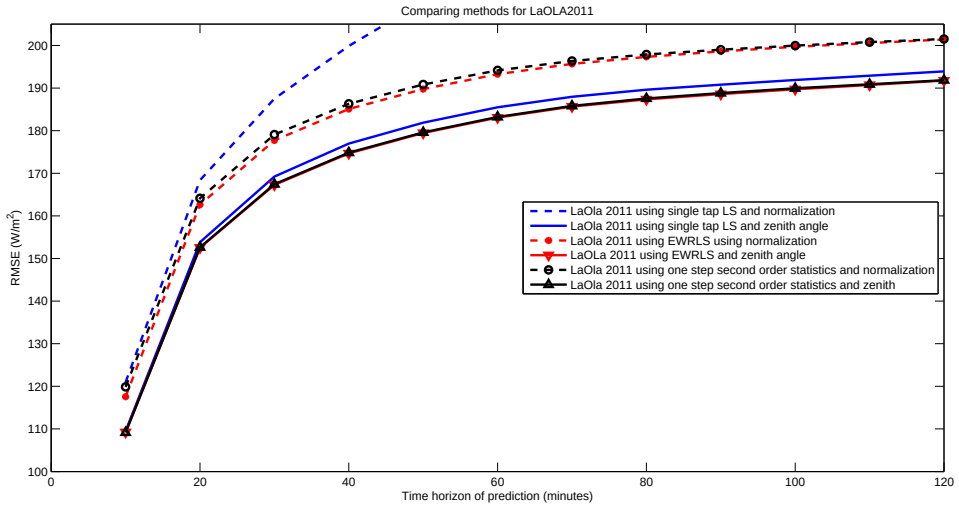

Figure 2.12: In simulation of LaOla using EWRLS and one step second order statistics with zenith angle give the best result

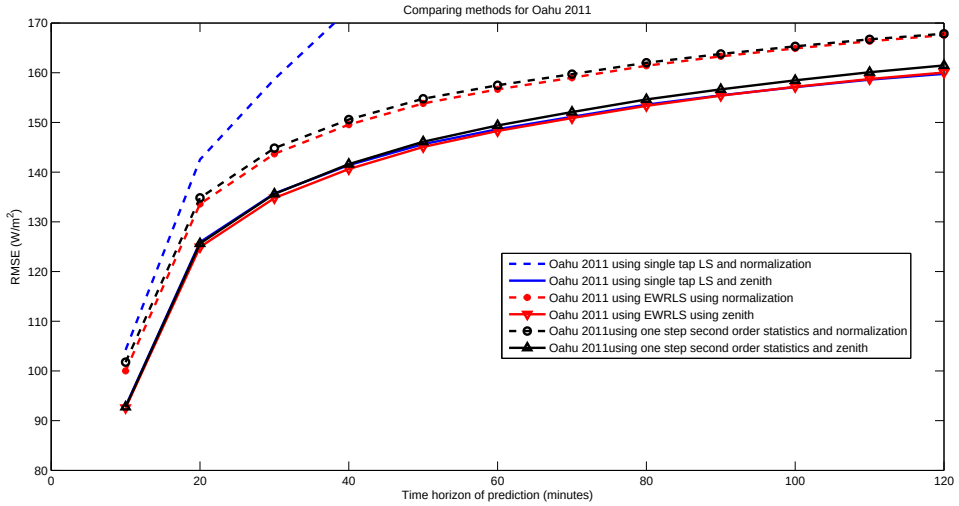

Figure 2.13: In Oahu station best result comes from LS with zenith and EWRLS with zenith angle

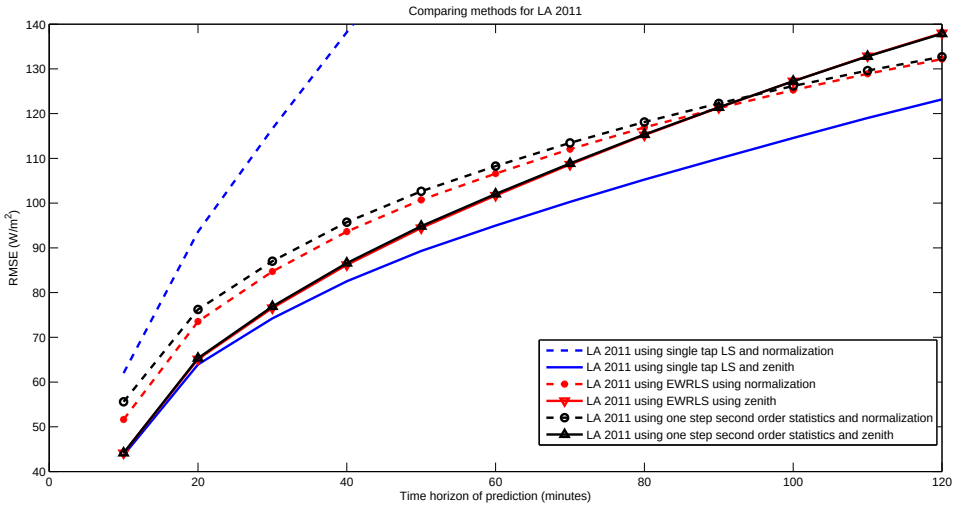

Figure 2.14: In LA station the best result comes from LS using zenith angle

2.5 Conclusion

In this chapter we implemented and evaluate simple forecasting methods for predicting solar radiation. We used two normalization methods for removing effects of time of day and day of the year. In the first method the solar radiation decoupled from time by dividing by corresponding cosine of zenith angle then LS, EWRLS and second order statistics are used as forecasting methods. In the second method we use 60 previous data and calculate mean and variance at each time of the day and use them to normalize the data to become zero mean and unit variance then forecasting methods like LS, EWRLS and second order statistics are used.

Simulation results for three different sites showed that dividing solar radiation by corresponding cosine of zenith angle is an effective way for removing both seasonal and daily effects.

Chapter 3

Prdiction Under Asymmetric Cost Functions

3.1 Introduction

Continuous balancing of load and generation is necessary for the electric grid. Power system balancing authorities (BAs) at each hour estimate the loads and schedule operation of conventional power plants. Integration of renewable generation into the grid has been increasing in recent years. However, renewable sources such as solar and wind are intermittent. Since it takes time to start additional conventional power plants, grid operators must predict the intermittent generation as well as load and commit enough generation resources in advance. To mitigate forecast errors, system operators allocate operating reserves - equivalent to creating a downward-biased forecast of "firm" renewable generation forecast - to ensure that during operation, generation always meets load [41] [42].

In this context, underestimation means that true renewable generation during operation

time is more than what the BA forecast. In this case during operation, output from conventional power plants is decreased below their committed level, so that generation equals load [43]. Output of conventional power plants could be decreased by a certain amount (available downward reserves) and if the magnitude of underestimation is more than available downward reserves, the excess renewable generation would be curtailed.

On the other hand, overestimation means that true renewable generation during operation time is less than what the BA forecast. So the committed generation capacity is not enough to meet load. In market-based power systems, this can result in extremely high prices in the balancing market. If additional generation is not available, the misforecast may lead to an area control error (ACE), drawing unscheduled power from neighboring BAs. In extreme cases or on isolated power systems, the BA may need to disconnect firm loads. ACE or load shedding are very undesirable for the BA and/or customers [44]. These also correspond to high economic costs, which manifest as fines paid by the BA for violating reliability standards [45], or the loss by customers of a valuable resource[46][47].

Put simply, in the case of overestimation of renewable generation, the BA will encounter shortages of generation and even may be forced to shed loads; however, in the case of underestimation of renewable generation, they can curtail the excess generating capacity. So overestimation is a more serious - and more costly - error than underestimation.

Therefore the solar and wind generation forecast problem in the BA's view is not symmetric. Many forecast models are trained using symmetric cost functions, but this may give sub-optimal results if forecast errors will impose asymmetric costs when used in practice. In this case, a forecast model using an asymmetric cost function may be more desirable.

The International Energy Agency (IEA) recommends the use of symmetric metrics such as Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Kolmogorov Smirnoff

Integral (KSI) for evaluating renewable energy forecasts [48][49]. Many authors who study solar irradiation forecasting tune their methods to minimize the RMSE, Mean Absolute Error (MAE) or Mean Absolute Percentage Error (MAPE) [50] which apply a symmetric cost for both underestimation and overestimation.

Renewable energy forecast tools can be divided into two broad categories - those that are trained using historical data [50][51], and those that are not. We do not address non-training-based methods, which include numerical weather prediction (NWP) [52][53], satellite image processing [54], or sky image processing [55]. Instead, we show the value of using asymmetric cost functions when training the first category of forecast tools.

A diverse array of forecast methods use a training phase. These include machine learning algorithms [56][57][58], artificial neural networks [59][60][61], fuzzy systems [62], hybrid methods [63][64], auto-regressive models [65], auto-regressive moving average (ARMA) models [66], and auto-regressive integrated moving average (ARIMA) [67]. In addition, training-based techniques are sometimes used to supplement non-training-based methods [68]. Statistical, training-based methods are more popular than physics-based models when making forecasts for short time scales (0-3 hours)[50], as needed for spinning reserve allocation.

Each of the training-based methods discussed above usually uses a symmetric cost function during the training phase. However, in many practical problems in economics the cost function is asymmetric. For example in dam construction underestimation of peak water level is more serious than overestimation [69]. In estimation of average life of the components of a spaceship, overestimation is usually more serious than underestimation. In this study, overestimation of renewable generation is more serious than underestimation.

To model the asymmetry, we use the asymmetric cost functions convex piecewise linear

(CPWL) and linear-Exponential (LinEx); these cost functions have been used successfully for optimization problems in economics, statistics and energy [69][70][71][72]. The simplest form of CPWL cost function (which is called LinLin) consists of two connected line segments - a constant positive slope (constant per-unit cost) for overestimation and a lower negative slope for underestimation errors (Fig. 3.1). In the LinEx cost function the per unit cost of overestimation increases exponentially as the error increases, but the per unit cost is relatively constant for underestimation errors. We note that the true cost function for renewable energy forecast errors is complex and unique to each BA. However, CPWL and LinEx make good, simplified proxies when evaluating forecast systems generically or training them for a particular region. They also allow us to illustrate some general properties of forecast methods tuned for asymmetric rather than symmetric cost functions.

A key to this chapter is establishing a general framework where we can make fair comparisons between the performance of forecast methods using asymmetric cost functions and other approaches using symmetric cost functions for solar radiation forecasts. We show that a method which creates a biased forecast by optimizing directly against the asymmetric function gives a better result than adding an optimal post-bias to a forecast previously trained with a symmetric cost function. We compare both "directly biased" and "indirectly biased" forecasts to emphasize the importance of using a directly biased forecast in this problem.

Online algorithms are interesting because of their suitability for real time applications, tracking gradual changes in statistics of input and hardware configuration [73][74]. For this reason, we also implement online algorithms based on the stochastic gradient descent or the least mean square (LMS) algorithm.

The chapter is organized as follows. In Section 3.2, the forecast problem is formulated

as an optimization problem; a forecast model using solar zenith angle is discussed and the use of CPWL and LinEx cost functions are justified. Section 3.3 presents the solution formulation for both CPWL and LinEx cost functions. For each of the cost functions we compare an "indirectly biased" and "directly biased" training method. In the indirectly biased method, the model's main training phase uses a traditional, symmetric cost function, which results in an unbiased model. Then a constant bias value is added, to minimize the asymmetric cost function as far as possible (with no other change to the forecast model). In the directly biased method, we use an asymmetric cost function (CPWL or LinEx) during the forecast model's main training phase. We also discuss methods for finding optimal parameter values in each of these cases. Section 3.4 introduces an online forecast method under these asymmetric cost functions. Section 3.5 presents the simulation results and discussion. A summary of results and conclusion is given in Section 3.6.

3.2 Problem statement

This section is divided into three subsections. In subsection 3.2.1, the forecast problem is formulated as an optimization problem based on a weight function of past and present observations. In subsection 3.2.2, a weight function that uses the solar zenith angle is explained. In subsection 3.2.3, we justify the use of CPWL and LinEx asymmetric cost functions in the utility scheduling problem.

3.2.1 Optimization Problem

Our objective is to minimize expected cost by adjusting forecast hypothesis parameters. Let the actual solar radiation at time n be x_n and the corresponding forecast be $\hat{x}_n$. We are

interested in k step ahead forecast using a window of past observations,

$$X_n = [x_n, x_{n-1}, ..., x_{n-m+1}]^T \tag{3.1}$$

where m is the window size.

Let us assume that the k step ahead forecast is a function of past and present observations

$$\hat{x}_{n+k} = h(X_n) \tag{3.2}$$

Then the optimization problem is

$$\underset{h}{\text{Minimize}} \sum_{i=1}^{M} L(h(X_i) - x_{i+k}) \tag{3.3}$$

where L is the loss function (either CPWL or LinEx in this chapter) and M is the total number of samples.

3.2.2 Hypothesis Model

Many solar power forecast methods use a training stage to adapt the model to conditions in a particular locale. In this chapter, we use a simple training-based method introduced in [32] to illustrate the benefits of using an asymmetric cost function during the training stage. This simplifies the discussion, and results found with this model may be generalizable to other training-based models.

The model we use for illustration is based on an autoregressive approach using the cosine of the solar zenith angle as we discussed in previous chapter. For illustration, we use an autoregressive forecast method but rather than regressing directly on recent irradiance measurements, this method uses irradiance normalized by $\cos\theta_z$. This normalization accounts for much of the daily and seasonal variation in irradiance, so that the model mostly

predicts the effect of the atmosphere on irradiance [32]. Our hypothesis model h is a linear combination of past data converted to the time of prediction:

$$\hat{x}_{n+k} = (\alpha_0 + \frac{\alpha_1 x_n}{\cos\theta_z(n)} + ... + \frac{\alpha_m x_{n-m+1}}{\cos\theta_z(n-m+1)}) \cos\theta_z(n+k) \tag{3.4}$$

where $\theta_z(n)$ is solar zenith angle at time n and $\alpha_0, \alpha_1, ..., \alpha_m$ are the weight parameters.

3.2.3 The Cost Functions

If the BA ignores all intermittent generation (i.e. forecasts zero output from these sources) it will schedule enough operating reserves at all times. However, unused reserves cost about 20% of per unit price of energy (i.e. in Hawaii about \$0.05/kWh) [42]. So forecast of intermittent generation is useful to avoid that cost. On the other hand if intermittent generation is overestimated, the BA may run short of generation and be forced to draw unscheduled power from neighboring BAs, or eventually shed loads. The BAs view load shedding as very undesirable and expensive. The cost of unscheduled power transfers is difficult to assess. The value of lost load (VOLL) due to load shedding is reported to be around \$8/kWh to \$24/kWh [75][76][77]. As a starting point in this study, let us assume VOLL to be \$10/kWh.

To model asymmetric costs, Granger introduced the piecewise linear LinLin function and suggested a useful although sub-optimal way to consider asymmetry, by adding a constant bias value to the predictor [70]. This is analogous to the simple approach of discounting renewable energy forecasts by a fixed amount when allocating spinning reserves.

Let ϵ be the forecast error given by

$$\epsilon_{n+k} = \hat{x}_{n+k} - x_{n+k} \tag{3.5}$$

So overestimation, which means the predicted value exceeds the actual value, corresponds to a positive error and underestimation which means the forecasted value is less than actual generation, corresponds to a negative error. The asymmetric trade off between underestimation ($0.05/kWh unnecessary cost) and overestimation ($10/kWh penalty fee) lead to a LinLin loss function.

$$LinLin(\epsilon) = \begin{cases} C_1\epsilon & \text{if } \epsilon > 0, \quad C_1 \approx \$10/kWh \\ -C_2\epsilon & \text{if } \epsilon \leq 0, \quad C_2 \approx \$0.05/kWh \end{cases} \tag{3.6}$$

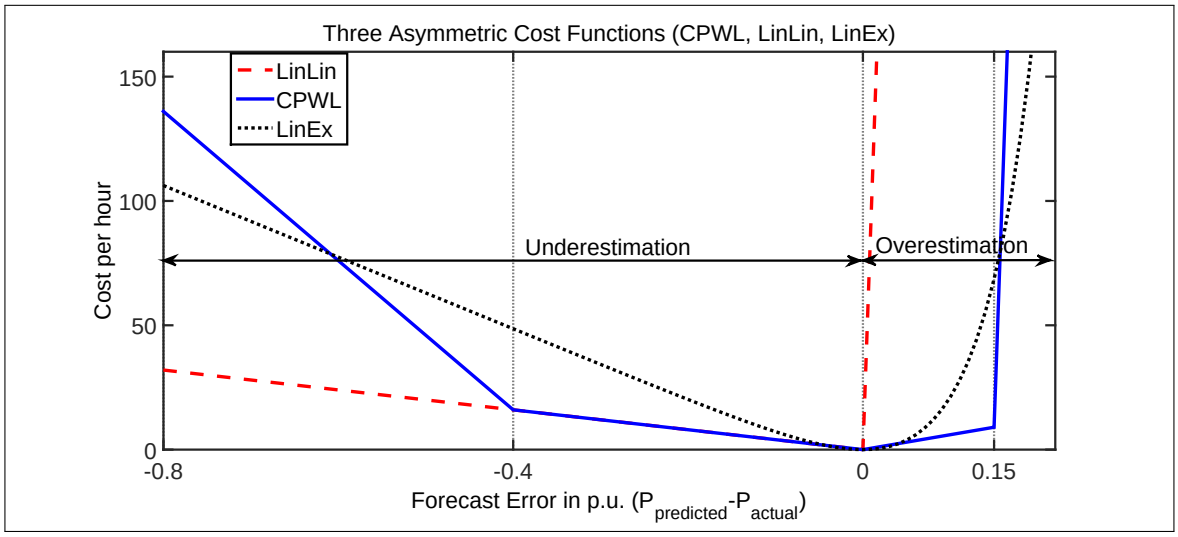

Figure 3.1: CPWL (C_1=-300/p.u., d_1=-0.4 p.u., C_2=-40/p.u., d_2=0, C_3=60/p.u., d_3=0.15 p.u., C_4=10000/p.u.), LinLin (C_1=-50/p.u.,C_2=10000/p.u) and LinEx (b=9, a=16/p.u.)

The LinLin loss function is the simplest asymmetric cost function we can use to distinguish between overestimation and underestimation. However, it is unable to represent cases where the per-unit cost increases as the magnitude of the error increases. A more complicated cost function could be obtained by using more line segments.

$$CPWL(\epsilon) = \max_{i=1,2,...p} C_i\epsilon + b_i \tag{3.7}$$

where p is number of line segments. Without loss of generality let us assume $C_1 < C_2 < ... < C_p$ and $d_1, d_2, ...d_{p-1}$ are intermediate points that determine the domain of each line segment. The solid line graph in Fig. 3.1 shows a CPWL cost function. It has four line segments with different slopes, the small positive slope shows the cost of upward reserves and the line with the steepest slope represents the penalty for errors which are larger than available upward reserves (high penalty for violating reliability standards and/or disconnecting firm load). Similarly, gentle negative slope shows the cost of downward reserves and the steeper negative slope represents cases where output of power plants could not decrease further (minimum generation level of power plants) and the excess renewable generation must be curtailed. An illustrative example in Fig. 3.1 slopes and break points are chosen by rule of thumb for a system with a single power plant which is set at 85% of rated power and must have minimum output power around 45% of rated power. The actual cost function may have more line segments representing the cost of different control actions which could be taken to mitigate errors and different break points based on power plant states.

Another popular asymmetric cost function is LinEx which was originally introduced for real estate assessment [78] and comprehensively discussed by Zellner [69]. The LinEx function could be a useful proxy for the costs of generation shortfall, which may rise as the size of error increases. These could occur if it costs more per MWh to correct a large error than a small one. For example, this could occur for renewable energy forecasts if a limited pool of reserves are used for renewable energy and other risks; small errors could easily be absorbed by the pool, but large errors could cause greater harm since increasingly expensive emergency resources may need to be brought online, or increasingly valuable loads may need to be shed. The reliability fines in the WECC region also depend in part on

the severity of violation [45]. Hence as a second case, we assume the load shedding cost is exponentially distributed among errors such that that small errors impose a smaller cost but large errors are more expensive. In this case we have the LinEx loss function given by

$$LinEx(\epsilon) = b(e^{a\epsilon} - a\epsilon - 1) \tag{3.8}$$

The a and b constants are called shape factor and scale factor respectively. In Fig. 3.1 the dotted line graph is a LinEx function and its shape and scale factor are selected to be similar to the given CPWL. Different values could be used to fit the function to any particular power system.

3.3 Solution Formulation

In a power system with asymmetric costs for forecast errors, the balancing authority will generally prefer to adopt a forecast method which is biased toward the lower-cost side. In this section we show two ways that a balancing authority could introduce this bias. We call these "indirectly biased" and "directly biased" forecasts. In later sections, we compare the effectiveness of these two biasing methods.

To create an indirectly biased forecast, a balancing authority would first use an unbiased forecast method, and then add a constant bias factor, with a value selected to minimize the asymmetric cost function. This approach could be used, for example, when a balancing authority adopts a solar forecast tool previously optimized to minimize mean squared error, and then allocates an additional reserve margin to compensate for possible forecast errors. This is equivalent to creating a downward-biased forecast of "firm" solar power[79]. Various methods that are commonly used create unbiased forecasts. In particular, any forecast

method trained to minimize squared error will produce an unbiased predictor, with a mean error equal to zero. Further, any forecast method with symmetrical error distributions will result in an unbiased predictor if it is trained with a symmetrical cost function.

To create a directly biased forecast instead of an indirectly biased forecast, the balancing authority would incorporate an asymmetric cost function directly into the training stage of the forecast tool, rather than adjusting the forecast after-the-fact.

This section is divided into two subsections. Subsection 3.3.1 is devoted to training or adjusting forecasts using CPWL cost functions and Subsection 3.3.2 uses LinEx cost functions. In both cases, we show techniques for optimizing indirectly and directly biased forecasts.

3.3.1 CPWL cost function

Indirectly biased forecast: Let the (unbiased) forecast error from the initial forecast model be ϵ and cumulative distribution function (CDF) of error be F_ϵ and probability density function of errors be $f(\epsilon)$. For CPWL cost function we have that

$$Loss_{total} = \int_{-\infty}^{+\infty} CPWL(\epsilon) f(\epsilon) d\epsilon \tag{3.9}$$

If we add a bias value β to the initial forecast, the cumulative loss with the CPWL cost function changes as there is a shift of β resulting in :

$$Loss_{total} = \int_{-\infty}^{+\infty} CPWL(\epsilon) f(\epsilon - \beta) d\epsilon \tag{3.10}$$

$$= \int_{-\infty}^{+\infty} CPWL(x + \beta) f(x) dx \tag{3.11}$$

To find the bias value β which minimizes cumulative loss, we have:

$$\frac{\partial Loss_{total}}{\partial \beta} = \int_{-\infty}^{+\infty} \frac{\partial CPWL(x+\beta)}{\partial \beta} f(x)dx = 0 \tag{3.12}$$

$$= \int_{-\infty}^{d_1-\beta} C_1 f(x)dx + \sum_{i=2}^{p-1} \int_{d_{i-1}-\beta}^{d_i-\beta} C_i f(x)dx + \int_{d_{p-1}-\beta}^{+\infty} C_p f(x)dx$$

$$= \sum_{i=1}^{p-1} (C_i - C_{i+1}) F(d_i - \beta) + C_p = 0 \tag{3.13}$$

Hence the optimal bias value to minimize CPWL cost will satisfy the following condition:

$$\sum_{i=1}^{p-1} (C_i - C_{i+1}) F(d_i - \beta^*) + C_p = 0 \tag{3.14}$$

The left side is nondecreasing in β^* so we can easily find the solution using a bisection method [80]. Hence, we find a best-possible indirectly biased solution as follows: first use an unbiased initial forecast method (here, we use least squares to find weight parameters of equation (3.4)); then calculate the distribution of errors and compute the optimal bias value that solve equation (3.14) using the bisection algorithm; finally add this bias value to the initial prediction.

Directly biased forecast: Our objective is

$$\min_{\alpha_0,\alpha_1,\ldots,\alpha_m} \sum_{n=1}^{M} CPWL(\hat{x}_{n+k} - x_{n+k}) \tag{3.15}$$

where $\hat{x}_{n+k}$ is computed using equation (3.4).

So we have

$$\min_{\alpha_0,\alpha_1,,\ldots,\alpha_m} \sum_{n=1}^{M} \max_{j=1,2,\ldots p} \{C_j[(\alpha_0 + \frac{\alpha_1 x_n}{\cos\theta_z(n)} + \ldots \tag{3.16}$$

$$+ \frac{\alpha_m x_{n-m+1}}{\cos\theta_z(n-m+1)}) \cos\theta_z(n+k) - x_{n+k}] + b_j\}$$

In order to eliminate the maximum segment, let us introduce new decision variables w_n for $n = 1, 2, 3, ..., M$ such that

$$w_n = \max_{j=1,2,...p} \{C_j[(\alpha_0 + \frac{\alpha_1 x_n}{\cos\theta_z(n)} + ... + \frac{\alpha_m x_{n-m+1}}{\cos\theta_z(n-m+1)})\cos\theta_z(n+k) - x_{n+k}] + b_j\} \tag{3.17}$$

So we have

$$\min_{\substack{w_1, w_2, ..., w_M \\ \alpha_0, \alpha_1, ..., \alpha_m}} \sum_{n=1}^{M} w_n \tag{3.18}$$

subject to

$$C_j[(\alpha_0 + \frac{\alpha_1 x_n}{\cos\theta_z(n)} + ... + \frac{\alpha_m x_{n-m+1}}{\cos\theta_z(n-m+1)})\cos\theta_z(n+k) - x_{n+k}] + b_j \leq w_n \tag{3.19}$$

for all $n = 1, 2, ..., M$ and $j = 1, 2, ..., p$

which is a linear programming problem.

3.3.2 LinEx cost function

Indirectly biased forecast: An indirectly biased solution could be found by using an unbiased forecast method (we used least squares to tune the parameters of equation 3.4) and then computing the optimal β. Again let our unbiased forecast error be ϵ and the probability density function of the errors be $f(\epsilon)$. If we add bias value β to the unbiased forecast, the errors are also increased by β so the cumulative loss with LinEx cost function becomes:

$$Loss_{total} = \int_{-\infty}^{+\infty} LinEx(\epsilon + \beta) f(\epsilon) d\epsilon \tag{3.20}$$

$$= b \int_{-\infty}^{+\infty} (e^{a(\epsilon+\beta)} - a(\epsilon + \beta) - 1) f(\epsilon) d\epsilon \tag{3.21}$$

To find the optimal bias value β which minimizes cumulative loss, we have:

$$\frac{\partial Loss_{total}}{\partial \beta} = abe^{a\beta} \int_{-\infty}^{+\infty} e^{a\epsilon} f(\epsilon) d\epsilon - ab \tag{3.22}$$

$$\Rightarrow \beta = -\frac{1}{a} \log(\int_{-\infty}^{+\infty} e^{a\epsilon} f(\epsilon) d\epsilon) \tag{3.23}$$

similar to Zellner's suggestion [69] :

$$\beta = -\frac{1}{a} \log(\mathrm{E}_{\epsilon}[e^{a\epsilon}]) \tag{3.24}$$

If there are enough samples available we can estimate β using the following:

$$\beta = -\frac{1}{a} \log(\mathrm{E}_{\epsilon}[e^{a\epsilon}]) \approx -\frac{1}{a} \log(\frac{1}{M} \sum_{i=1}^{M} e^{a\epsilon_i}) \tag{3.25}$$

where M is the total number of samples and ϵ_i is the error corresponding to ith sample.

Directly biased forecast: Here again we use a forecast method based on zenith angle as in equation (3.4); however for selection of parameters we consider a LinEx cost function. We want to adjust $\alpha_0, \alpha_1, ..., \alpha_m$ such that the following objective function will be minimized:

$$J = \sum_{n=1}^{M} LinEx(\hat{x}_{n+k} - x_{n+k}) \tag{3.26}$$

where $\hat{x}_{n+k}$ is computed using equation (3.4). Since this optimization does not have an analytical answer, we use the gradient descent algorithm. We note that LinEx is a convex function [69] so the gradient descent algorithm with proper step size converges to the global minimum. Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, ...\alpha_m]^T$; then the $\boldsymbol{\alpha}$ is iteratively updated by following equation.

$$\boldsymbol{\alpha}_{i+1} = \boldsymbol{\alpha}_i - \eta \nabla J \tag{3.27}$$

where η is step size and ∇J is the gradient vector, computed by the following equations:

$$\nabla J = [\frac{\partial J}{\partial \alpha_0}, \frac{\partial J}{\partial \alpha_1}, ..., \frac{\partial J}{\partial \alpha_m}]^T \tag{3.28}$$

$$\frac{\partial J}{\partial \alpha_0} = ab \sum_{n=1}^{M} [\cos\theta_z(n+k)(e^{a(\hat{x}_{n+k}-x_{n+k})}-1)] \tag{3.29}$$

similarly for $j = 0, 1, 2, ..., m-1$

$$\frac{\partial J}{\partial \alpha_{j+1}} = ab \sum_{n=1}^{M} [\frac{x_{n-j}\cos\theta_z(n+k)}{\cos\theta_z(n-j)}(e^{a(\hat{x}_{n+k}-x_{n+k})}-1)] \tag{3.30}$$

.

3.4 Online Methods

In many instances information about data is not complete and data statistics may be nonstationary. In these cases online learning algorithms which continually update their weights often give superior performance over the batch algorithms [81] we formulated in Section 3.3. The algorithmic simplicity of online algorithms is also an important issue when we deal with large scale problems [82] especially in real time. Online algorithms are also more preferable in term of hardware implementation due to high modularity [73].

Therefore in this section we implement online forecast methods based on the stochastic gradient descent algorithm. The first subsection is devoted to online formulation for a CPWL cost function and the second subsection is dedicated to the LinEx cost function.

3.4.1 CPWL Cost Function

Similar to least mean squares (LMS) algorithm [74] which uses the instantaneous estimate of gradient vector for squared error cost, we use an instantaneous estimate of the gradient vector for the CPWL cost function.

$$\hat{J} = CPWL(\hat{x}_{n+k} - x_{n+k}) \tag{3.31}$$

where $\hat{x}_{n+k}$ computed using equation (3.4). The instantaneous estimate of gradient vector computed by the following equations.

$$\nabla \hat{J} = [\frac{\partial \hat{J}}{\partial \alpha_0}, \frac{\partial \hat{J}}{\partial \alpha_1}, ..., \frac{\partial \hat{J}}{\partial \alpha_m}]^T \tag{3.32}$$

To compute the gradient, define function $g(x)$ using the following equation

$$g(x) = \begin{cases} C_1 & \text{if } x < d_1 \\ ... & \\ C_i & \text{if } d_{i-1} \leq x < d_i. \\ ... & \\ C_p & \text{if } d_{p-1} \leq x \end{cases} \tag{3.33}$$

The partial derivatives are calculated using

$$\frac{\partial \hat{J}}{\partial \alpha_0} = \cos\theta_z(n+k) g(\hat{x}_{n+k} - x_{n+k}). \tag{3.34}$$

In the same way, for $j = 0, 1, 2, ..., m-1$,

$$\frac{\partial \hat{J}}{\partial \alpha_{j+1}} = \frac{x_{n-j}\cos\theta_z(n+k)}{\cos\theta_z(n-j)} g(\hat{x}_{n+k} - x_{n+k}). \tag{3.35}$$

Let $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, ...\alpha_m]^T$. Then $\boldsymbol{\alpha}$ is iteratively updated by following equation.

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n - \eta \nabla \hat{J} \tag{3.36}$$

The total cost up to time $J(n)$ is the cumulative sum of instantaneous cost, $\hat{J}$, So

$$J(n) = J(n-1) + \hat{J}(n) \tag{3.37}$$

3.4.2 LinEx Cost Function

Similarly, instantaneous LinEx cost is given by

$$\hat{J} = LinEx(\hat{x}_{n+k} - x_{n+k}) \tag{3.38}$$

where $\hat{x}_{n+k}$ computed using equation (3.4). The instantaneous estimate of the gradient vector is computed by the following equations.

$$\nabla \hat{J} = [\frac{\partial \hat{J}}{\partial \alpha_0}, \frac{\partial \hat{J}}{\partial \alpha_1}, ..., \frac{\partial \hat{J}}{\partial \alpha_m}]^T \tag{3.39}$$

$$\frac{\partial \hat{J}}{\partial \alpha_0} = ab[\cos\theta_z(n+k)(e^{a(\hat{x}_{n+k}-x_{n+k})}-1)] \tag{3.40}$$

Similarly for $j = 0, 1, 2, ..., m-1$,

$$\frac{\partial \hat{J}}{\partial \alpha_{j+1}} = ab[\frac{x_{n-j}\cos\theta_z(n+k)}{\cos\theta_z(n-j)}(e^{a(\hat{x}_{n+k}-x_{n+k})}-1)] \tag{3.41}$$

The $\boldsymbol{\alpha}$ is iteratively updated by the following equation.

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n - \eta \nabla \hat{J} \tag{3.42}$$

Adding a momentum term to the learning rule could increase the learning rate; however, a constant momentum factor (γ) results in oscillation in the learning curve [83]. For this reason we tried decreasing the momentum factor such that at initial iterations the momentum factor is high and leads to faster learning but in later iterations the momentum factor decreases to zero to avoid over learning. We found the following time varying momentum term worked well

$$\gamma_n = \frac{\gamma_0}{(1+\frac{n}{N})} \tag{3.43}$$

where N is the number of samples per year.

The learning algorithm with a momentum term is implemented using the following equations:

$$\boldsymbol{\Delta\alpha}_{n+1} = \gamma_n \boldsymbol{\Delta\alpha}_n - \eta \nabla \hat{J} \tag{3.44}$$

$$\boldsymbol{\alpha}_{n+1} = \boldsymbol{\alpha}_n + \boldsymbol{\Delta\alpha}_{n+1}$$

3.5 Simulation Results

For simulation, we retrieved solar irradiance data for three sites from a National Renewable Energy Lab (NREL) website, www.nrel.gov/midc/. The names and details of the sites are shown in Table 3.1. In order to concentrate on hours of the day with significant solar radiation, we removed night hours and only considered nine hours per day. To compute generated power from solar radiation and meteorological information (temperature, wind speed), we first estimate solar module temperature using equation(3.45) [84].

$$T_{mod} = T_{amb} + I\frac{T_{NOCT} - T_0}{I_0 + h(\nu - \nu_0)(T_{NOCT} - T_0)} \tag{3.45}$$

where $T_0 = 25°C$ is reference temperature, $\nu_0 = 1m/s$ is reference wind velocity, $I_0 = 800W/m^2$ is reference irradiation, $h = 6.62W/°Cm^2$ is a convection parameter, T_{NOCT} is the temperature of the module under reference conditions. T_{abm}, ν and I are ambient temperature, wind speed and solar radiation. Then we derate the nominal conversion coefficient according to the maximum power temperature coefficient of the solar module ($\approx -0.44\%/°C$) as given in the data sheet for Sharp ND-R250 module.

In order to get comparable results, we report per unit cost for each forecast method, which is the ratio of the annual cost when using that forecast method to the annual cost if no forecast is used (renewable generation assumed to be zero).

3.5.1 Results of Batch Methods

To compare the benefit of forecasts using both indirectly biased forecast and directly biased forecasts, we used one year of data for training and the next year for testing. In each case, we prepared hour-ahead forecasts, using one to nine taps (number of weight

Table 3.1: The names and details of the sites used for simulations

Sites	Loactions	Recored Time	Resolution
Hawaii La Ola Lanai	Latitude:20.76685 N Longitude:156.92291 W Elevation: 381 meters AMSL	12/14/2009 to 12/14/2012	1 minute
North Carolina Elizabeth City	Latitude:36.28 N Longitude:76.22 W Elevation: 26 meters AMSL	1/1/2005 to 12/31/2013	5 minutes
California Los Angeles	Latitude:33.966674 N Longitude:118.42282 W 27 meters AMSL	1/1/2011 to 12/31/2014	1 minute

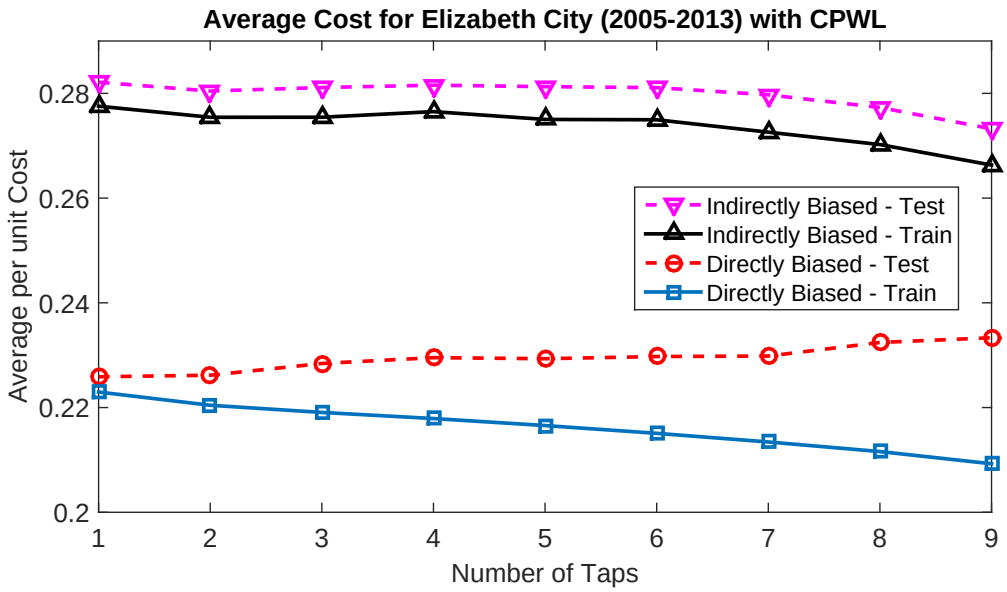

Figure 3.2: With the CPWL cost function, the directly biased (linear programming) method is more efficient than the indirectly biased forecast. If the original unbiased forecast is used, the per unit cost becomes around 2 which is worse than the trivial no forecast cost.

parameters in equation (3.4)). The per unit cost for the CPWL cost function is shown in Fig. 3.2 and for the LinEx cost function is shown in Fig. 3.4. As is clear from the figures, in both cases the cost for directly biased forecasts is significantly less than from indirectly biased forecasts.

The graphs in Fig. 3.2 are the result of averaging per unit annual cost of 2005-2013 of the Elizabeth City data set using the CPWL cost function. Although increasing the number of taps reduces the training cost for the direct method (linear programming), the cost for test data slightly increases for more than one tap due to over-fitting. So by using the direct approach we decrease the cost to 22% of the no-forecast cost, which is less than the 27% achieved by an indirectly biased forecast.

In Fig. 3.3 both the distribution of errors and the CPWL cost function are shown (the

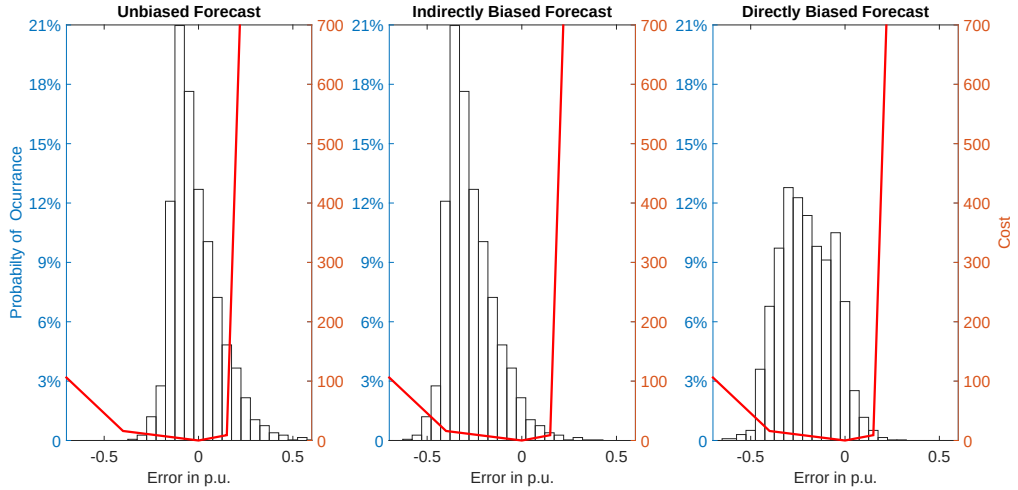

Figure 3.3: Histograms of forecast errors for three different scenarios for the CPWL cost function (unbiased, indirectly biased, and directly biased using linear programming)

total in each case is the product of the probability and the cost). For unbiased forecast we have many positive errors as well as negative errors since there is no cost difference between positive and negative errors in the training phase. Most errors are close to 0. By adding bias (negative number) to the unbiased forecast, the distribution of errors is shifted to the left since cost for large overestimations is huge. Now, most of errors are close to -0.25. On the other hand, the directly biased method shifts errors to the left so that a very small portion of errors are in the steep penalty zone. We also note that the directly biased method also has fewer large-magnitude negative errors than the indirectly biased forecast method.

Three sets of graphs are shown in Fig. 3.4 to show per unit cost in 2005-2013 for Elizabeth City using a LinEx cost function. The dotted line graphs are for the case that one year is used for training and another year used for test. Training costs decrease by increasing the number of taps while the test costs increase for more than one tap. The

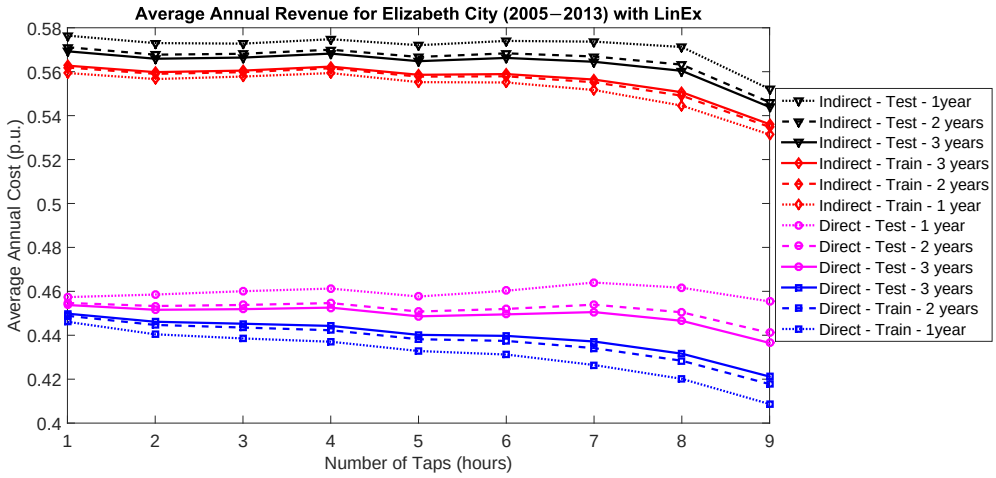

Figure 3.4: The direct biased forecast method has more benefit than indirectly biased forecast. A more complex model gives better performance but needs more training data (LinEx cost function)

dashed and solid line graphs respectively show the cases for two and three years of data used for training. In both direct and indirect methods the use of more training data decreases the difference between test and training and the use of a more complex model increases risk of over fitting.

By using a cross validation technique, we use eight years for training and one other year for testing. The results of the cross validation are shown in Fig. 3.5. In this way there is fair agreement between training and test. We have achieved about 43% of the no-forecast cost using a directly biased forecast, which is significantly less than the 54% achieved by an indirectly biased forecast. In Fig. 3.6 the distribution of errors for an unbiased forecast is given and most of the errors are close to 0. By adding bias (negative number) to the unbiased forecast the total distribution is shifted to the left since underestimation has less cost than overestimation. Now, most of the errors are close to 0.25 due to the bias. On

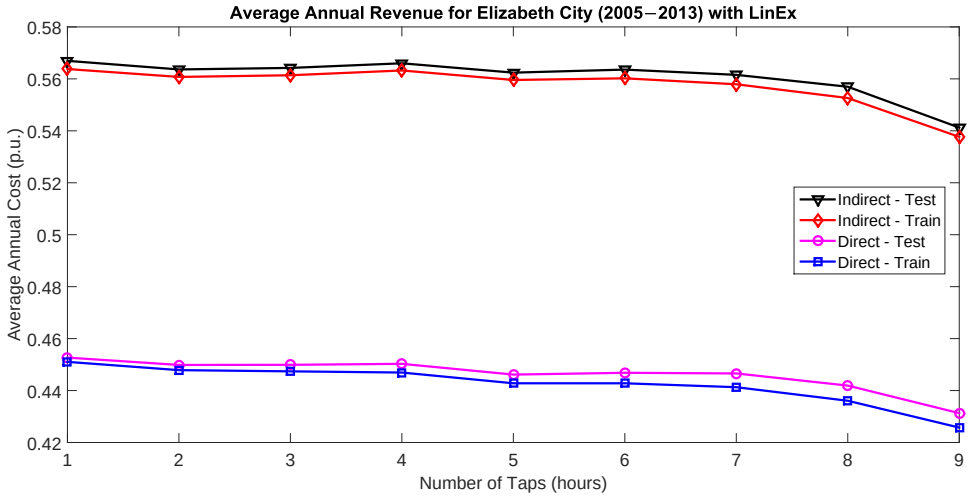

Figure 3.5: More training data gives better agreement between training and testing results (nine fold cross validation)

the other hand, the directly biased forecast shifts errors to the left so that a small portion of errors are positive. This is achieved with the asymmetric cost function that severely penalizes positive errors. We again note that the directly biased method also has fewer large-magnitude negative errors than the indirectly biased forecast method. Fig. 3.7 shows that the directly biased method performs much better than the indirectly biased method as the shape factor (asymmetry) of the LinEx cost function increases, although costs increases for both methods.

3.5.2 Results of Online Methods

In the steepest descent algorithm, several iterations are required to reach the optimal weight vectors and in each iteration all training samples are used; so each sample is used several times. In order to use each sample multiple times in the online method, we use a resampling technique i.e. we use nine yearly datasets of Elizabeth City seven times, then

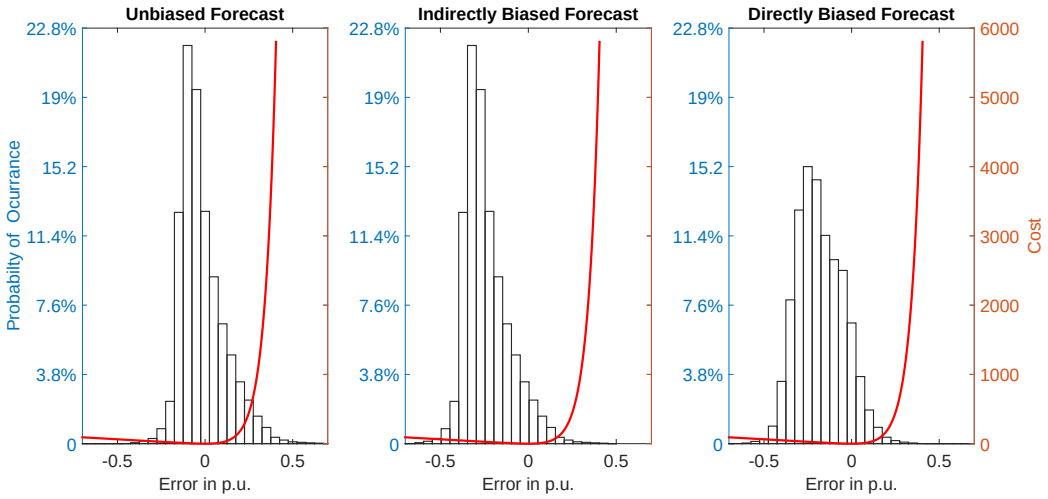

Figure 3.6: Histograms of forecast errors for three different scenarios for the LinEx cost function (unbiased, indirectly biased, and directly biased forecast)

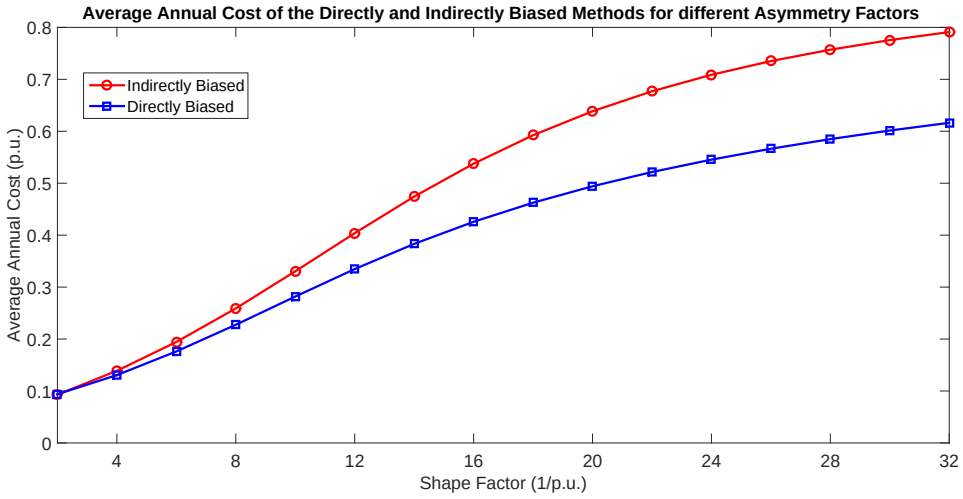

Figure 3.7: As the shape factor (asymmetry) in the LinEx cost function increases the directly biased method performs much better than the indirectly biased method.

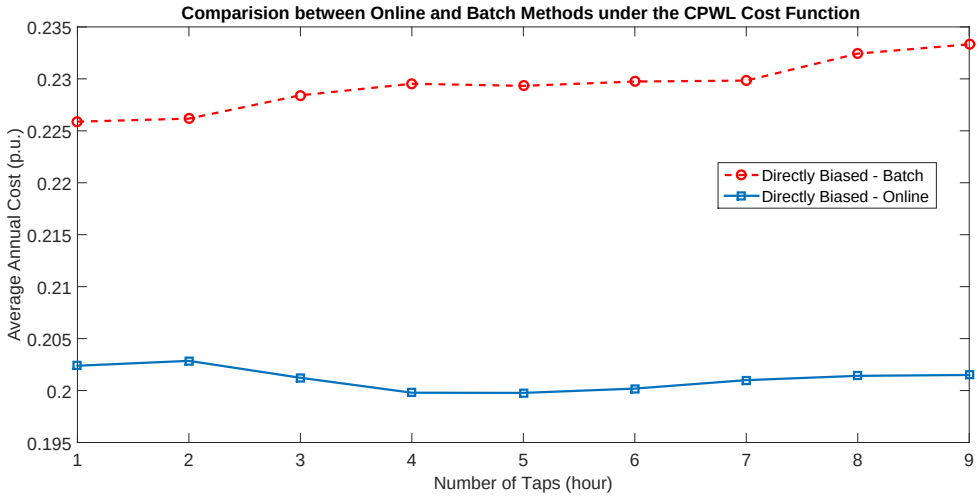

Figure 3.8: The per unit cost of online method for CPWL cost function is slightly less than corresponding batch method due to tracking ability of the online method.

the total 63 yearly datasets are randomly ordered and used as input to the online learning algorithm.

Fig. 3.8 shows the per unit average annual cost for the CPWL cost function using the online forecast method with different taps. It reveals that increasing the number of taps beyond four does not give any performance improvement. The per unit cost is 20% for four taps, which is 2% less than the 22% test cost of the corresponding batch method. This slight improvement arises from the tracking ability of the online method.

Online algorithms are interesting especially for the cases where there is no recorded data and the operator is just starting to record. In these cases, the online algorithm at each iteration receives a new data sample and improves its model. Fig. 3.9 shows the learning curves of the online method with one and four taps under a CPWL cost function; at initial iterations the cost is around 100% since data is not enough for good forecast. As more data are obtained the forecast model is tuned better, so the cost decreases over time. The four

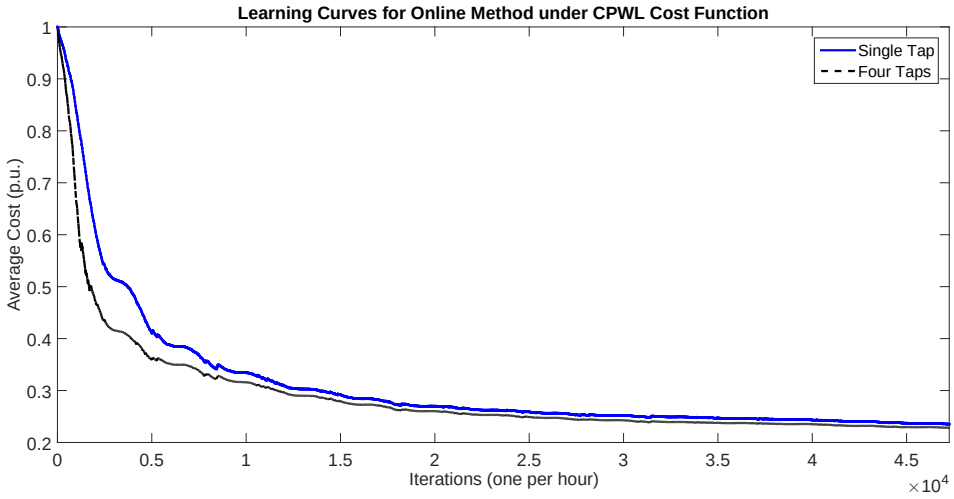

Figure 3.9: Learning curves for methods with different taps under CPWL cost function

tap method learns faster, so it gives slightly better results.

As shown in Fig. 3.10 the per unit average annual cost for the LinEx cost function using the online forecast method is around 40% which is 3% better than the 43% for the corresponding batch methods. Also, the number of taps does not have a significant effect.

The learning curves of the online method with different tap numbers under the LinEx cost function are shown in Fig. 3.11. While there is not a significant difference in final performance, the method which uses more taps learns faster than others.

We repeated the same simulations for the datasets for Hawaii and Los Angeles with three years of data. For all of these datasets the savings using directly biased methods are substantially more than the indirect method as shown in Table 3.2.

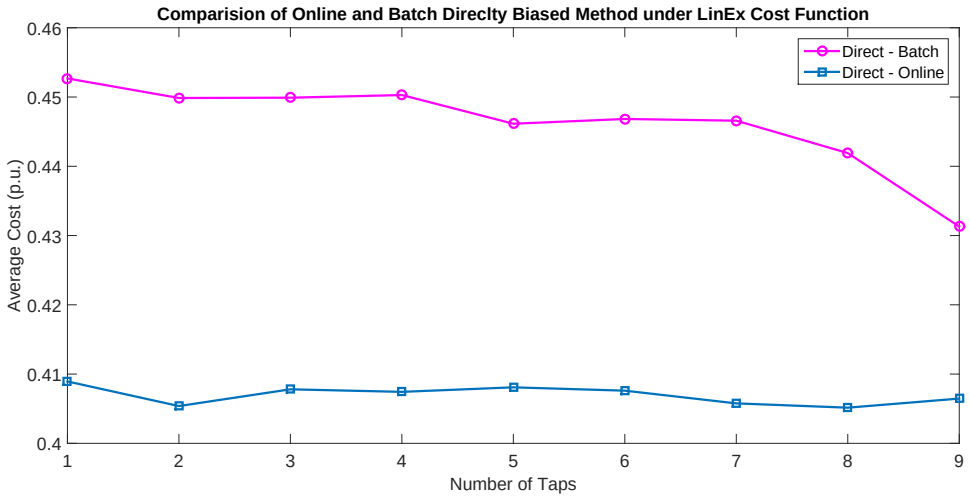

Figure 3.10: The per unit cost of online method for the LinEx cost function is less than corresponding batch method. Also, increasing number of taps does not have significant effect.

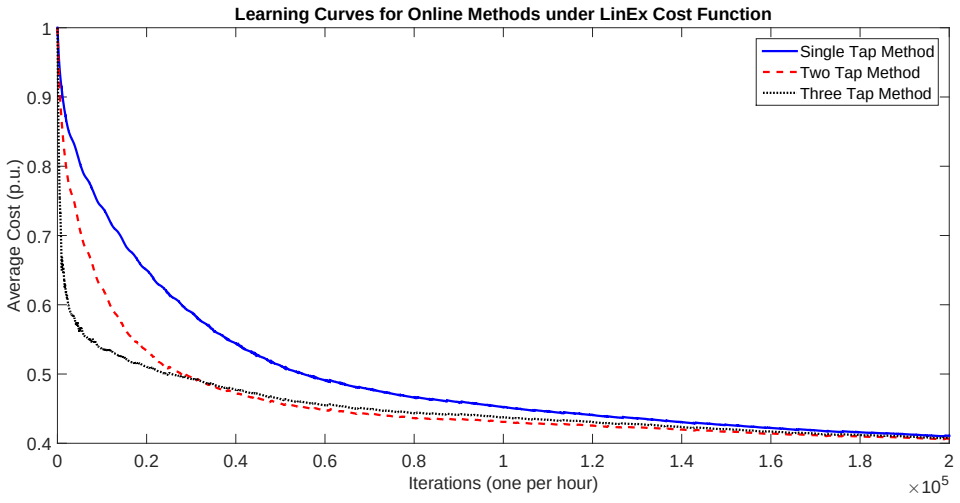

Figure 3.11: Under LinEx cost function, the method which has more taps learn faster than others, however, the final performance is similar to others.

Table 3.2: Average annual per unit cost for datasets using different methods

Cost Function	CPWL			LinEx		
Method	Batch		Online	Batch		Online
	Indirect	Direct	Direct	Indirect	Direct	Direct
Elizabeth City	27%	22%	20%	54%	43%	40%
Hawaii	36%	26%	25%	57%	44%	42%
Los Angeles	17%	14%	14%	38%	31%	31%

3.6 Conclusion

While many researchers have studied the problem of forecasting solar radiation, they usually evaluate their methods using symmetric criteria like root mean square error (RMSE) or mean absolute error (MAE). However, grid operators have more concern about shortage of production rather than its abundance, i.e. overestimation of resources has more serious consequences than underestimation. So in the BA's view the cost function is not symmetric. For this reason we discussed solar radiation forecast under CPWL and LinEx as asymmetric cost functions which are better fitted to the grid operator problem. For each of these cost functions we used two scenarios i.e. adding bias to an unbiased forecast or formulating a directly biased forecast which considers the cost function from the outset. Under CPWL cost the forecast is formulated as a linear program and for LinEx cost we formulated the problem as a convex optimization and solved it by a gradient descent algorithm. Our simulations show that directly biased forecasts have a significant advantage. Simulation results also show that this difference becomes greater as asymmetry in the cost function increases.

We also implemented the least mean square (LMS) algorithm according to CPWL and LinEx cost functions to create an online method. Using a decaying momentum term in the learning rule increases the rate of learning. The proposed online method gives an improvement over batch solutions due to better tracking ability.

We have shown the necessity of using asymmetric cost functions directly in the training phase of simple autoregressive forecast models. More sophisticated learning algorithms based on other methods such as neural networks or wavelets might potentially give better performance, and use of asymmetric cost function when training these models needs further investigation.

Point forecasts only give a single value for prediction of a random variable; on the other hand, probabilistic forecasts give more information using a probability distribution over future quantity (for example one CDF function for a random variable). When a probabilistic forecast is given, for each prediction we can solve equation (3.14) for the CPWL cost function or use equation (3.25) for the LinEx case to compute an optimal value for that prediction.

Here we discussed forecast methods for single sites using their past solar radiation observations. Forecast methods for multiple sites and incorporating exogenous data like weather forecasts are also interesting and left for future research.

Chapter 4

Probablistic Forecast of Solar Irradiation

4.1 Introduction

Traditionally, solar forecasting methods have a single output value for each prediction which is usually called a point forecast. The performance of point forecasts are usually evaluated based on a symmetric cost function of the errors (the difference between predicted and observed values). For example root mean square (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) are commonly used [50]. Many researchers proposed point forecast for solar irradiation prediction which includes machine learning algorithms [56–58], artificial neural networks [59–61], fuzzy systems [62], hybrid methods [63, 64], auto-regressive models [65], auto-regressive moving average (ARMA) models [66], and auto-regressive integrated moving average (ARIMA) [67]. Point forecasts can also be used when the cost function is asymmetric and approaches for direct predic-

tion under asymmetric cost functions like convex piecewise linear, LinLin and LinEx are proposed in [33–35].

However, for many decision makers point forecasts that only give a single value for prediction of a random variable is not enough and more information about future quantity is required. Probabilistic forecasts can give a more complete stochastic characterization using probability distribution over future quantity (for example predicting a CDF function for a random variable). Considering the limitations of point forecasts, a few papers have recently proposed probabilistic forecast for solar radiation prediction. For example, [7] suggests a Bayesian method to estimate parameters of the predictive distribution which was assumed to be a modified Gamma distribution. Three different stochastic differential equation models are proposed by [8]. These models are first fitted to a training data set and subsequently evaluated on a one-year test set. [9] uses a quantile regression forest to obtain a set of quantile values from observed solar power and corresponding numerical weather prediction. An analog ensemble method proposed in [10] is based on a set of historical numerical weather forecasts and corresponding observed solar power. A combination autoregressive moving average (ARMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models are used in [11] to obtain probabilistic forecast of solar irradiance from historical observations. Extreme learning machines (ELM) are used for regression purposes and are trained to obtain nonparametric probabilistic forecast by predicting quantiles [12]. Quantile regression is used in [13] as a competitive indirect method for solar forecasting under asymmetric cost function. However, probabilistic solar forecast is still immature and requires more in depth research [14, 15].

This chapter proposes two parametric probabilistic forecast methods using beta and power distribution and improves their performance metrics by a combining procedure based

on beta transformed linear opinion pool. The proposed approach can be used to extend the results of many different point forecast methods which are minimizing RMSE or MSE.

The rest of the chapter is organized as follows. The probabilistic forecast and evaluation metrics are described in Section 4.2. Methods for obtaining probabilistic forecast using beta and power distribution as well as the procedure of combining forecasts are given in Section 4.3. In Section 4.4 simulation results are shown and discussed. A summary of results and conclusion is given in Section 4.5.

4.2 Problem Statement

4.2.1 Probabilistic Forecast

The objective of a probabilistic forecast is to find a predictive distribution to predict stochastic characterization of response variable when information about explanatory variables is given. When the response variable is a binary event the conditional probability of success given the explanatory variable is a probabilistic forecast. Logistic regression is a method that is commonly used for probabilistic forecast of the binary events. When the response variable is a continues random variable (which solar irradiation is), a conditional pdf can describe complete stochastic characterization. Let $(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), ..., (y_n, \boldsymbol{x}_n)$ be training observations where $\boldsymbol{x}_i$ is explanatory variable (in general a vector), y_i is response variable and n is number of observations. The goal is to use training data and obtain the conditional pdf $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ or equivalently predictive CDF :

$$F_{Y|\boldsymbol{x}}(y) = \mathbb{P}(Y \leq y|X = \boldsymbol{x}) = \int_{-\infty}^{y} f_{Y|\boldsymbol{X}}(u|\boldsymbol{x})du \tag{4.1}$$

There are nonparametric and parametric approaches for obtaining probabilistic forecast. Nonparametric methods do not make any assumptions about shape of the density and concentrate on finding a set of quantile values [13, 85]. On the other hand, parametric methods assume a parametric distribution (like truncated Gaussian or Weibull [86]) in order to estimate predictive distribution.

4.2.2 Evaluation of Probalistic Forecasts

In point forecasts there are metrics like RMSE or MAE which are used to evaluate performance of different methods on train and test set. Similarly, metrics are required to evaluate performance of probabilistic forecasts. In this section some scoring rules are explained.

Logarithmic Score is a local scoring rule used to evaluate performance of a predictive pdf on a realized observation. Let f be the predictive pdf and x be an observation the logarithmic score is given by

$$LogS(f, x) = -\log(f(x)) \tag{4.2}$$

In fact, it is negative of log likelihood function and the smaller score is better performance. Average logarithmic score can be used for multiple observations and multiple predictive pdfs, which is again related to negative of log likelihood when observations are independent.

Continuous Ranked Probability Score is one of the most widely used accuracy metrics for probabilistic forecasts. Let F be predictive CDF and after the fact observation x is realized. The CRPS given by :

$$CRPS(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{I}(y - x))^2 dy \tag{4.3}$$

Where $\mathbb{I}$ is the step function i.e. equals zero when $y < x$ and equals one when $y \geq x$. The notion behind CRPS is integration of squared difference between predictive CDF and CDF of the best predictor which determines the outcome surely i.e. a CDF that jumps form zero to one at the value that will be observed later.

The CRPS is a generalization of the mean absolute error (MAE) for the case of probabilistic forecasts. It has been showed that equation (4.3) is equivalent to : [87]

$$CRPS(F, x) = \mathbb{E}_F|X - x| - \frac{1}{2}\mathbb{E}_F|X - X^*| \tag{4.4}$$

where X and X^* are two independent copies from distribution F with finite mean. The unit for CRPS is same as unit for the observation.

Calibration means statistical consistency between forecast and realized observations. Probability integral transform histogram and empirical distribution are two useful tools for evaluating calibration.

For a CDF-valued random quantity F and the observation Y, the PIT of probabilistic Forecast F is given by

$$Z_F = F(Y_-) + (F(Y) - F(Y_-))V \tag{4.5}$$

where V is a standard uniform random variable that is independent of F and for all $y \in \mathbb{R}$, $F(y_-) = \lim_{x \uparrow y} F(x)$ [87]. If F is continuous, the probability integral transform (PIT) is the random variable :

$$Z_F = F(Y) \tag{4.6}$$

Let F be CDF-valued random quantity with PIT Z_F:

Forecast F is marginally calibrated if $\mathbb{E}[F(y)] = \mathbb{P}(Y \leq y)$ for all $y \in \mathbb{R}$.

Forecast F is probabilistically calibrated if PIT of Z_F has standard uniform distribution.

A forecast that is ideal relative to some information is both marginally calibrated and probabilistically calibrated[87].

Calibration can be quantified by Kolmogorov-Smirnov test which is used to compare an empirical distribution with a reference distribution (one sample KS) or to compare two empirical distributions. So the one sample KS is used to test if PIT has uniform distribution and two sample KS is used to determine if average of predictive CDF (first empirical distribution) and marginal distribution of samples have same underlying distribution.

Probabilistic calibration can be quantified by variance of the PIT values. For the ideal calibrated forecast, PIT has standard uniform distribution so the variance is $1/12$. The forecast is called under-dispersed when the variance of PIT is more than $1/12$ and over-dispersed if the variance of PIT is less than $1/12$. The PIT histogram of an under-dispersed forecast is U shape which means it predicts extreme probabilities more often than it should. For an over-dispersed the PIT histogram is inverse U shape which means central probabilities are predicted more often than they should.

Sharpness is concentration of a probabilistic forecast. The goal of probabilistic forecast is to maximize its sharpness while maintaining calibration metrics at acceptable level. Sharpness is usually reported by average width of a specific prediction intervals for example average difference between 25% and 75% quantiles or average difference between 5% and 95% quantiles.

4.3 Methods

4.3.1 Using Beta Distribution

We can assume that each predictive CDF is a beta distribution. Beta distribution is chosen since solar irradiation is continuous and restricted between zero and clear sky irradiation (which can be normalized to the interval (0,1)). It also can effectively model the density of different sky conditions by choosing appropriate shape factors. In Fig 4.1 probability density function (pdf) of different shape factors are shown. The green curve is for situation that a bright sunny day with a small possibility of clouds is predicted. On the other hand the purple curve predicts a thick cloud with small chances for clear sky. Beta distribution is mostly a unimodal density and only can model two modes if the modes are located at extreme points but is still useful for our application. The pdf for a beta distribution is given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad x \in (0,1) \tag{4.7}$$

where $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$.

While Maximum likelihood estimation of beta parameters requires solving nonlinear equations using numerical methods or approximating solution, the parameters can be estimated using method of moments[88]. Let $\bar{x}$ be the sample mean and $\bar{v}$ be the sample variance. The estimated parameters are given by :

$$if \quad \bar{v} < \bar{x}(1-\bar{x}) \tag{4.8}$$

$$\hat{\alpha} = \bar{x}(\frac{\bar{x}(1-\bar{x})}{\bar{v}} - 1) \tag{4.9}$$

$$\hat{\beta} = (1-\bar{x})(\frac{\bar{x}(1-\bar{x})}{\bar{v}} - 1) \tag{4.10}$$

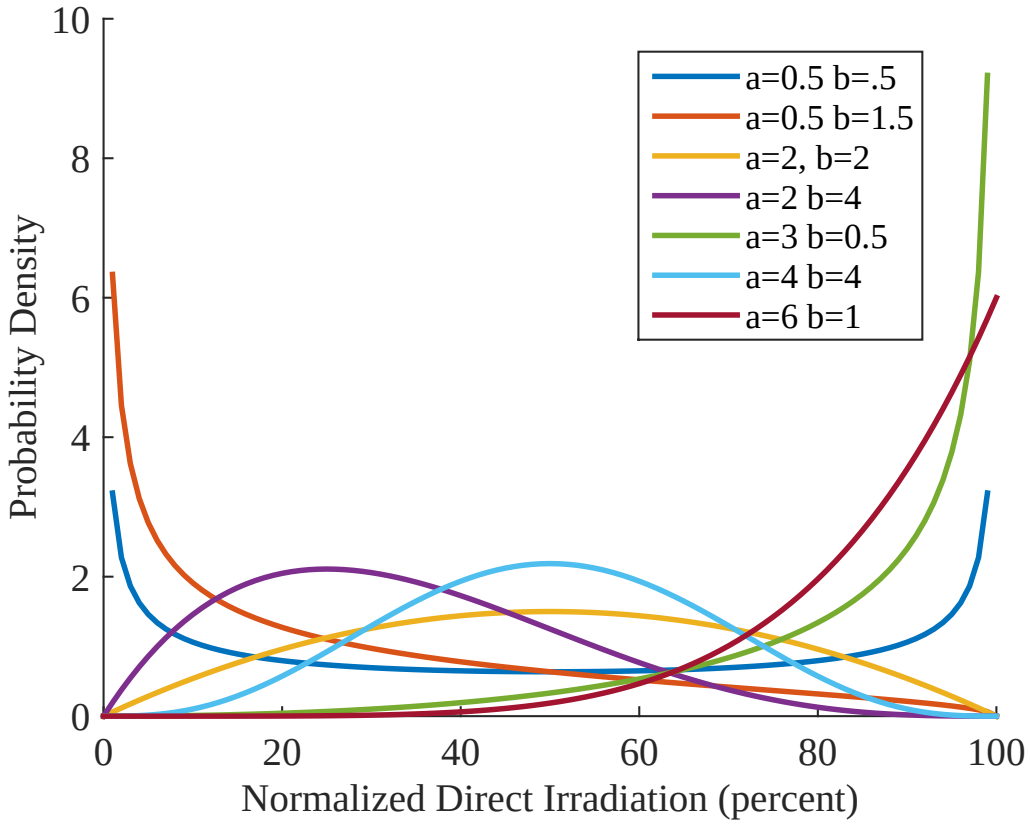

Figure 4.1: Beta distribution with different shape parameters

Equation (4.8) must be satisfied in order to get a valid solution ($\hat{\alpha} > 0$ and $\hat{\beta} > 0$).

Hence the probabilistic forecast using beta distribution can be reduced to estimation of conditional mean and conditional variance. Estimation of conditional mean is straightforward using LS technique. We also use LS to directly find the conditional variance, however this method is tricky as direct calculation of conditional variance sometimes lead to a negative number. Therefore we adjust the conditional variance to a constant minimum value if the method gives an invalid conditional variance.

4.3.2 Using Two Sided Power Distribution

Standard two sided power distribution is a generalized version of triangular distribution which is defined by following probability density function on interval (0,1)

$$f(x) = \begin{cases} k(\frac{x}{c})^{k-1} & 0 < x \leq c \\ k(\frac{(1-x)}{(1-c)})^{k-1} & c < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

where c is mode of the distribution and $k > 0$ is the order. This is designed to have the simplicity of a triangular distribution but also allow thiner tails than the standard triangular distribution. For a given c and k, the mean of the distribution is given by

$$m_X = \frac{(k-1)c + 1}{k+1} \tag{4.12}$$

A few examples of probability density function for this distribution are shown in Fig. 4.2. Similar to the beta distribution, power distribution can model different sky conditions.

Let $X_1, X_2, ...X_n$ be iid observations from a standard two sided power distribution. The procedure for estimation of parameters using maximum likelihood as well as method of moment is well explained in [89]. In the maximum likelihood method, first the mode is estimated by sorting observation and finding the observation that maximize the likelihood function without concern about the order and in the second step order parameter is estimated based on the estimated mode. However, this procedure is not applicable to our problem as we can not sort observations conditionally.

The method of moment uses sample mean and sample variance for making a third order polynomial equation. Positive and real roots of this third order equation are potential solution for the order parameter; the mode parameter for the potential solutions is computed

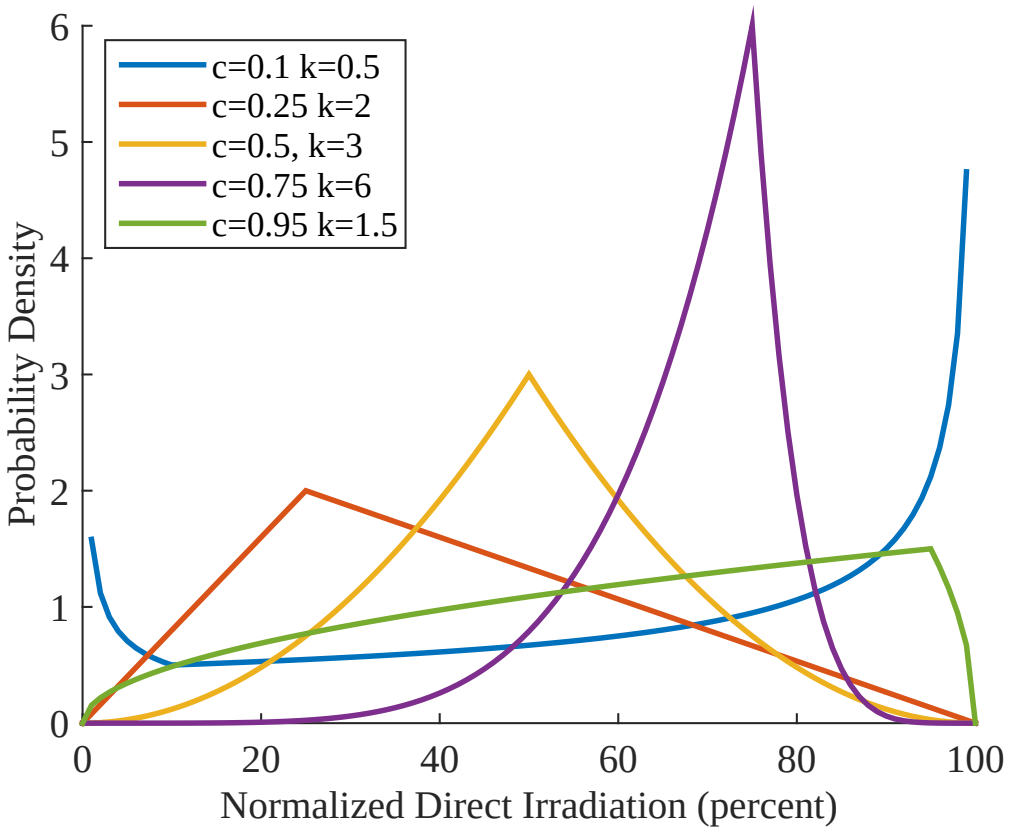

Figure 4.2: Two sided power distribution for different darameters

by obtaining inverse of equation (4.12) and finding the valid solution which leads to the mode parameter in interval (0,1). This procedure is applicable to probabilistic forecast problem as we can find the conditional mean and conditional variance then follow this method to find the parameters of predictive distribution. However, there will be systematic dependency between results of power distribution and the results of the beta distribution method as they use the same input. For example, if there is an error in estimation of conditional variance both methods will pass the error to their output and therefore the error also remains in the combined result.

In order to boost improvement when two methods are combined we propose to estimate

mode and order of the distribution (i.e. c, k), using an iterative method which combines the method of moments and maximum likelihood. For a given c_i, we find the order parameter which maximizes log likelihood function (which is concave and differentiable with respect to k) by taking derivative and set it equal to zero. We obtain

$$k_i = \frac{-n}{\sum_{X_j<c_i} \log(\frac{X_j}{c_i}) + \sum_{X_j>c_i} \log(\frac{1-X_j}{1-c_i})} \tag{4.13}$$

where X_j are observations and i is the iteration number. For the obtained k_i, we use second step of the method of moments which estimate the mode by obtaining inverse of equation (4.12) and bounding the results between zero and one.

$$c_{i+1} = \max(0, \min(1, \frac{(k_i+1)\bar{x}-1}{k_i-1})) \tag{4.14}$$

where $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} X_j$ is the sample mean. Then we again tune the order parameter in next iteration. After few iterations the equations approaches its fixed point i.e. the solution. We may have convergence issue when k is close to one but this is less of a concern as $k = 1$ resembles a standard uniform distribution which does not have a unique mode.

Hence, to obtain a probabilistic forecast, we first estimate the conditional mean using LS then above method is used to compute conditional mode and the order of the predictive distribution.

4.3.3 Combining Probabilistic Forecasts

In many cases, there are several probabilistic forecasts that use dependent or independent information and each of them has its advantages. It is useful to combine the forecasts and get one with superior performance. The most popular aggregation method is the linear pool [90]. The linear pool is successfully applied in many applications. However, it has

been shown that the output of a linear pool is more dispersed than the most dispersed input forecast [91, 92]. Therefore, it is good to aggregate under-dispersed forecasts to obtain a more calibrated forecast; however, if several calibrated forecast are aggregated by a linear pool, the output is uncalibrated.

Gneiting and Ranjan proposed to use a beta-transformed linear pool to combine multiple probabilistic forecast [91]. The method adds a mapping using a beta distribution to the traditional linear pool. Let $F_1, F_2, ...F_m$ be m different CDF-valued probabilistic forecast. The CDF of aggregated forecast is given by :

$$G(y) = B_{\alpha,\beta}(\sum_{k=1}^{m} w_k F_k(y)) \tag{4.15}$$

where $B_{\alpha,\beta}$ is the CDF of beta distribution with α and β parameters and w_k are nonnegative weights which add up to one. α and β parameters and w_k weights are optimized to minimize the average logarithmic score of the aggregated forecast on training data. We can use the same method for improving an uncalibrated forecast which is a special case of when there is only one input.

Here, there are two probabilistic forecasts, F_1 is CDF-valued forecast using beta distribution and F_2 is CDF-valued forecast using standard power distribution. We find three parameters α, β and w_1 (We note that $w_2 = 1 - w_1$.) that minimize logarithmic score of the combined forecast by using Newton-Raphson method.

4.4 Simulation Results and Discussion

For simulation, we again use the same datasets of chapter 3. For each of the sites one year is used for training and another year is used for test. We have tried different possible

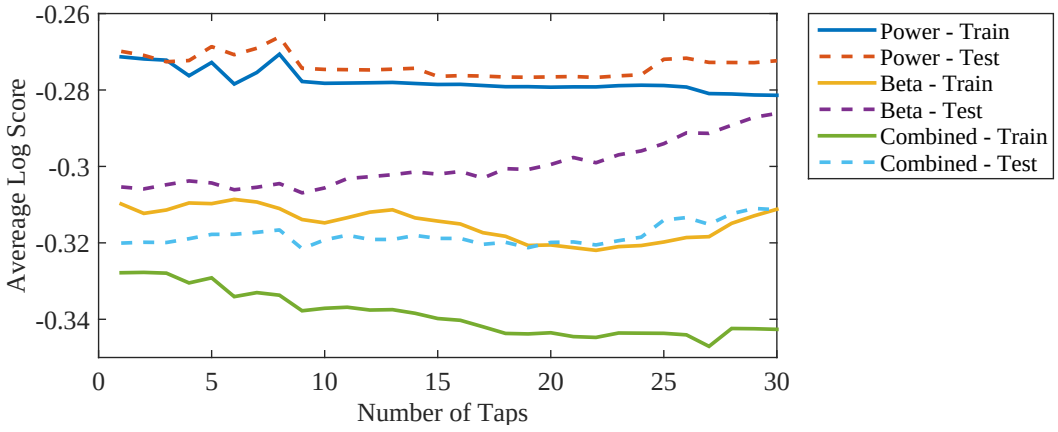

Figure 4.3: Performance in term of logarithmic score in LaOla

combinations based on the dataset size and the average performance is reported here. In order to make the comparison easier we report the normalized result (so the data is first normalized using zenith angle [32] and then scaled to the interval (0,1)) unless otherwise stated.

Fig.4.3, Fig.4.4 and Fig.4.5 respectively show the average logarithmic score of probabilistic forecasts for LaOla, Los Angeles and Elizabeth City. Solid lines show the training score whereas the test score are shown by dashed line. The beta method has lower logarithmic score i.e. predicts better than power method and the combined forecast is the best. We note that the training logarithmic scores do not monotonically decrease as number of taps increases. This is justified since both beta and power method use LS to find conditional mean by tuning the weight of each tap. However, more taps tend to indirectly decrease the training logarithmic score. The test logarithmic score may increase due to over-fitting.

Fig.4.6, Fig.4.7 and Fig.4.8 show the average CRPS as a function of number of taps. There is fair agreement between test and training results. The training score decreases when more taps are used. There are clear decreases for 9 tap, 18 tap, and 27 tap preditors that

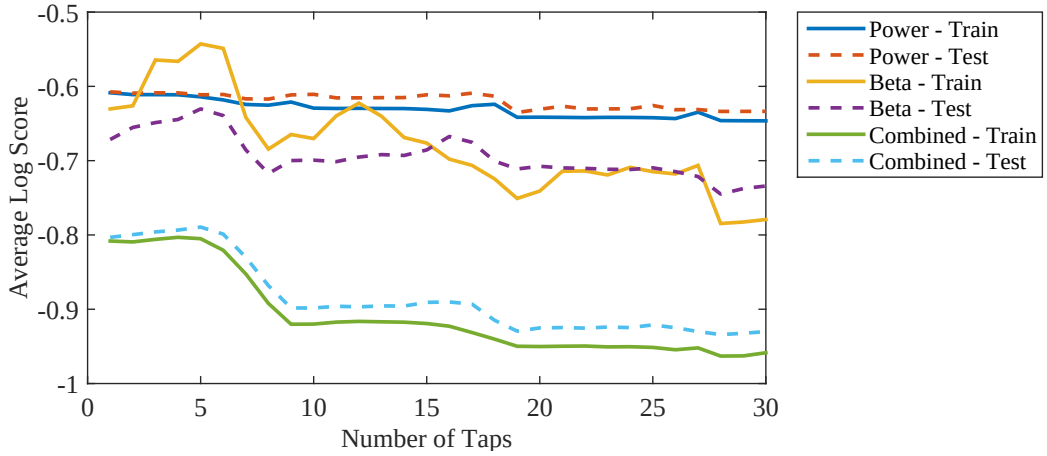

Figure 4.4: Performance in term of logarithmic score in Los Angeles

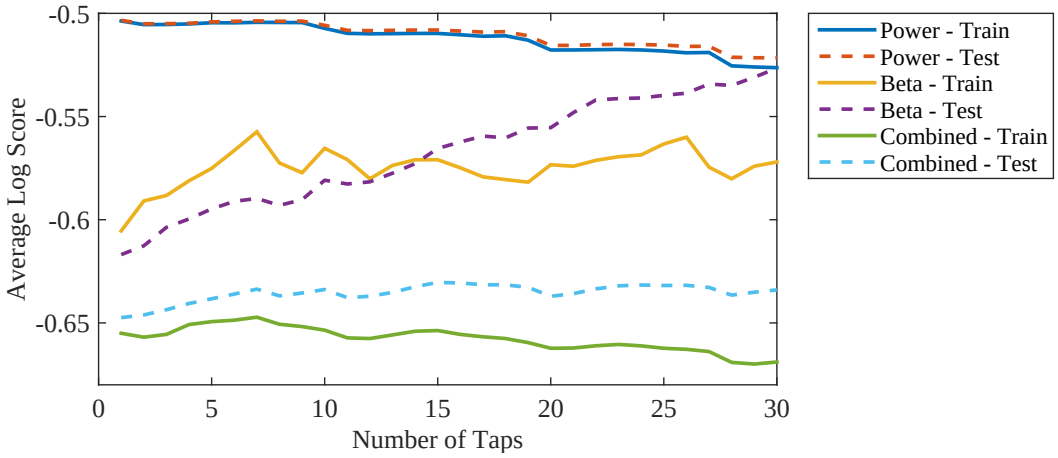

Figure 4.5: Performance in term of logarithmic score in Elizabeth City

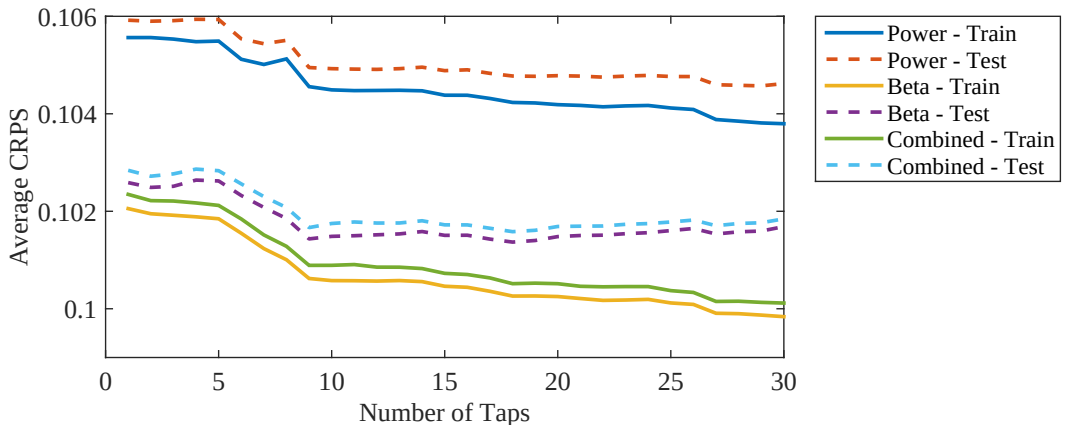

Figure 4.6: Performance in term of average CRPS in LaOla

shows that the same time of the day in past days has more information than other times. In terms of CPRS, the beta method has advantages over power method and combining them marginally improves the CRPS in Los Angeles and Elizabeth City but marginally increases CRPS in LaOla dataset.

Fig.4.9 shows marginal calibration of LaOla which reveals the good marginal calibration as the maximum CDF difference of the methods is small enough to pass the two sample KS test at 1% confidence. In Fig.4.10 marginal calibration of Los Angeles is shown. While combined forecast and beta method are significantly better calibrated than the power one, none of the methods can pass the KS test and therefore all of them are not marginally calibrated.

In Fig.4.11 marginal calibration of Elizabeth City is shown. While both combined forecast and beta method are calibrated and pass KS test at 5% confidence, the power method is not marginally calibrated.

Fig.4.12 shows probabilistic calibration of LaOla. The probabilistic calibration tends to improve with higher number of taps. Since maximum CDF difference of between PIT

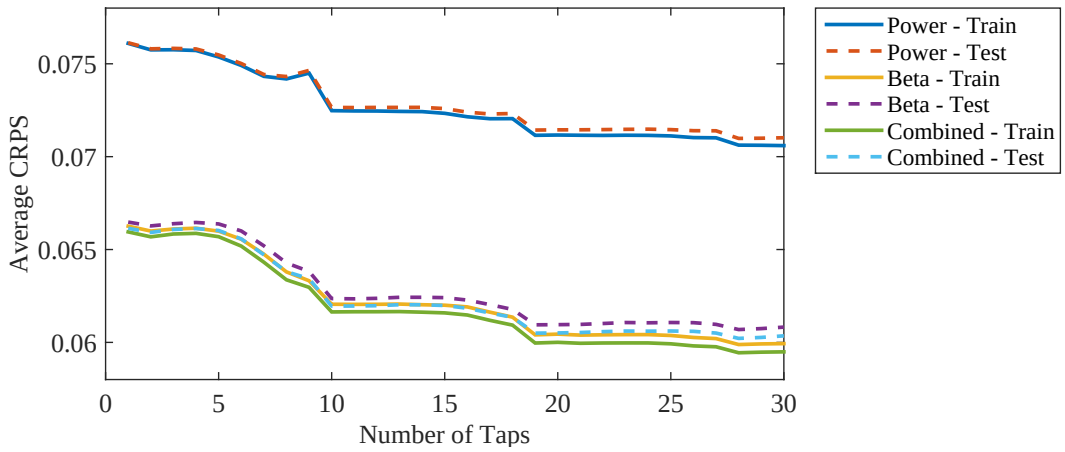

Figure 4.7: Performance in term of average CRPS in Los Angeles

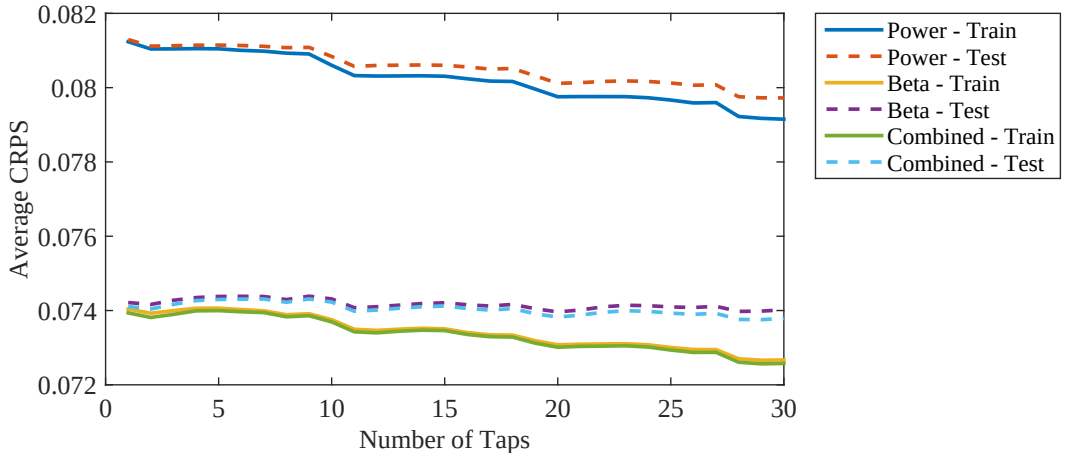

Figure 4.8: Performance in term of average CRPS in Elizabeth City

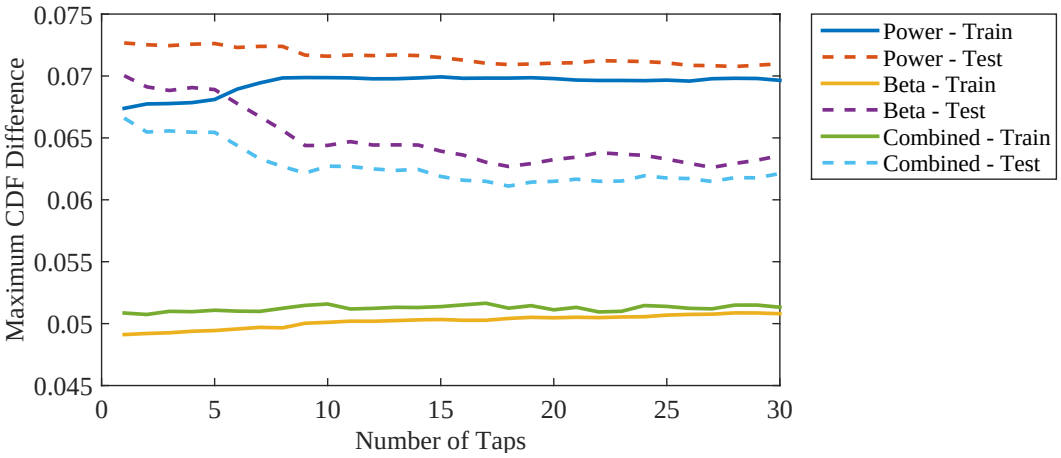

Figure 4.9: Performance in term of marginal calibration in LaOla

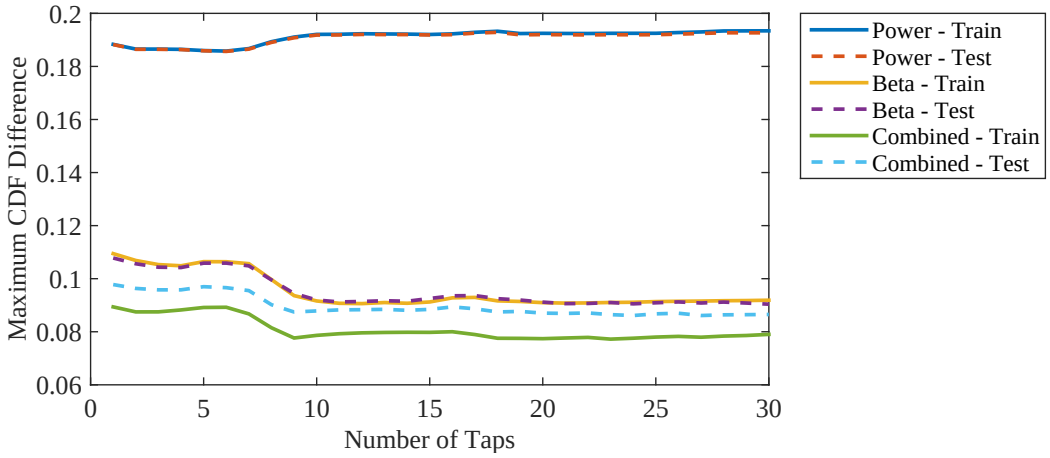

Figure 4.10: Performance in term of marginal calibration in Los Angles

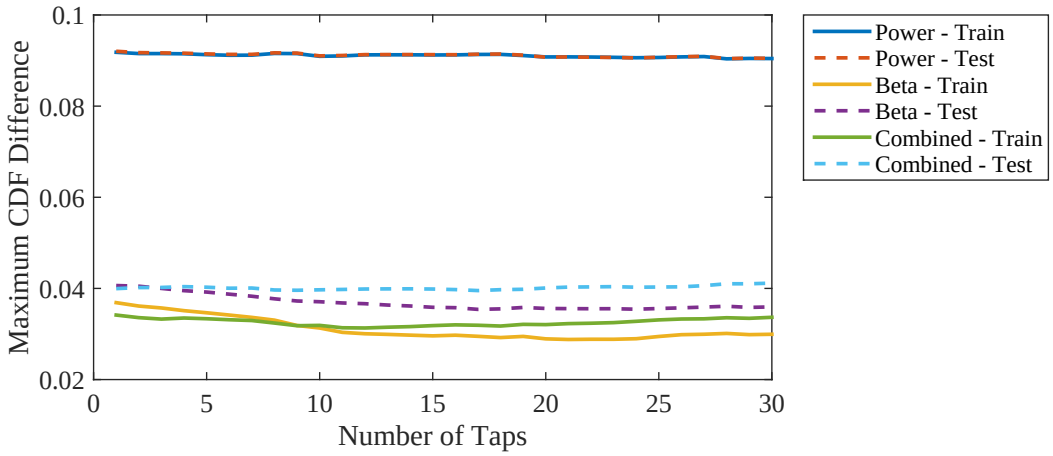

Figure 4.11: Performance in term of marginal calibration in Elizabeth City

and uniform distribution one sample KS test must be used i.e. a lower critical value with respect to marginal calibration. The beta method has better calibration and with more than 17 taps can pass the one sample KS test at 5%confidence whereas others can only pass the test at 1% confidence.

Fig.4.13 shows probabilistic calibration for Los Angeles. The combined method gives the best calibration; however, none of the methods can pass the one sample KS because of large difference between the CDF between PIT and uniform distribution. The probabilistic calibration for Elizabeth City is shown Fig.4.14. The combined method gives the best calibration and pass the one sample KS but other two methods can not pass it.

The PIT histograms for LaOLa, Elizabeth City and Los Angeles are shown in Fig. 4.15, Fig. 4.16 and 4.17. The PIT for LaOLa is very close to uniform distribution so the forecast is probabilistically calibrated. While the PIT for Elizabeth city has more density in the center. This inverse-U shape histogram indicates that the forecast is over-dispersed. The probabilistic forecast for Los Angeles is the most over-dispersed.

A forecast is marginally calibrated if expected predictive CDF converges to the true

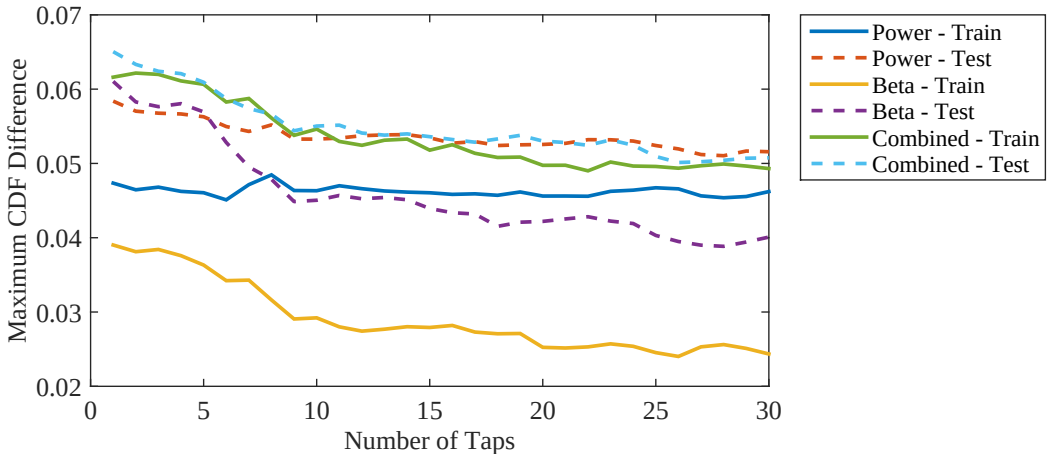

Figure 4.12: Performance in term of probabilistic calibration in LaOla

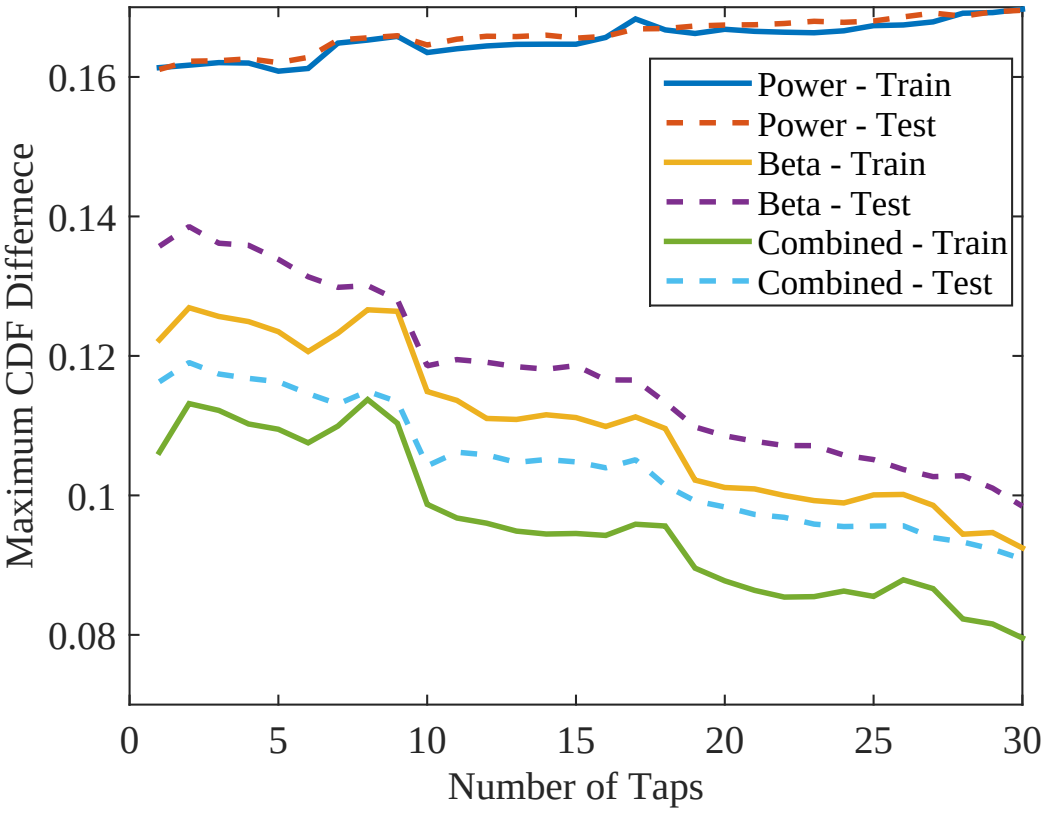

Figure 4.13: Performance in term of probabilistic calibration in Los Angeles

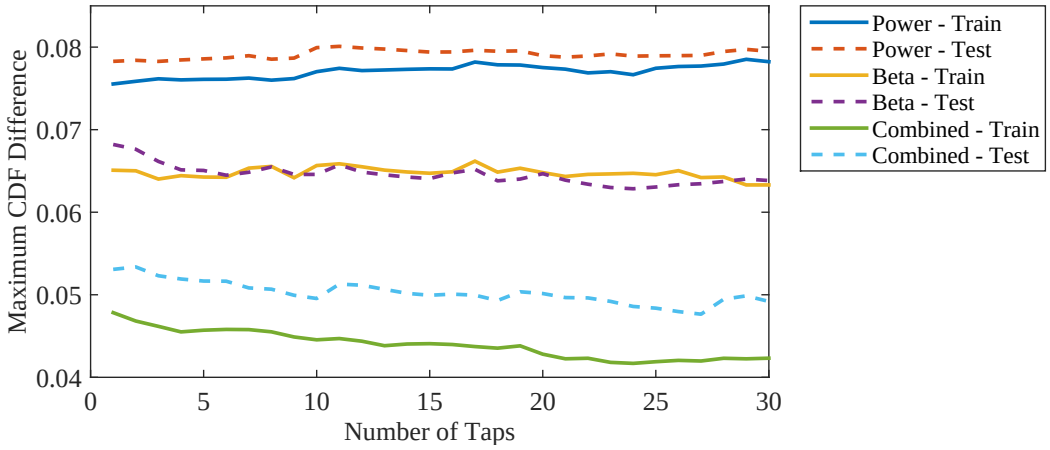

Figure 4.14: Performance in term of probabilistic calibration in Elizabeth City

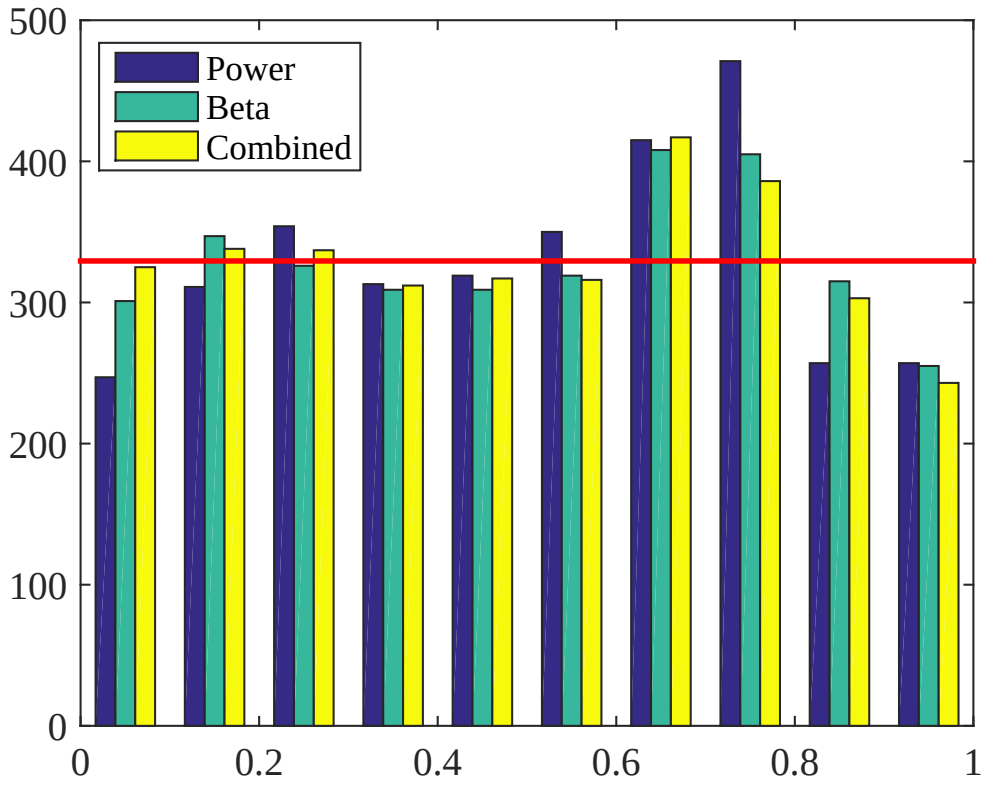

Figure 4.15: PIT histogram for LaOla

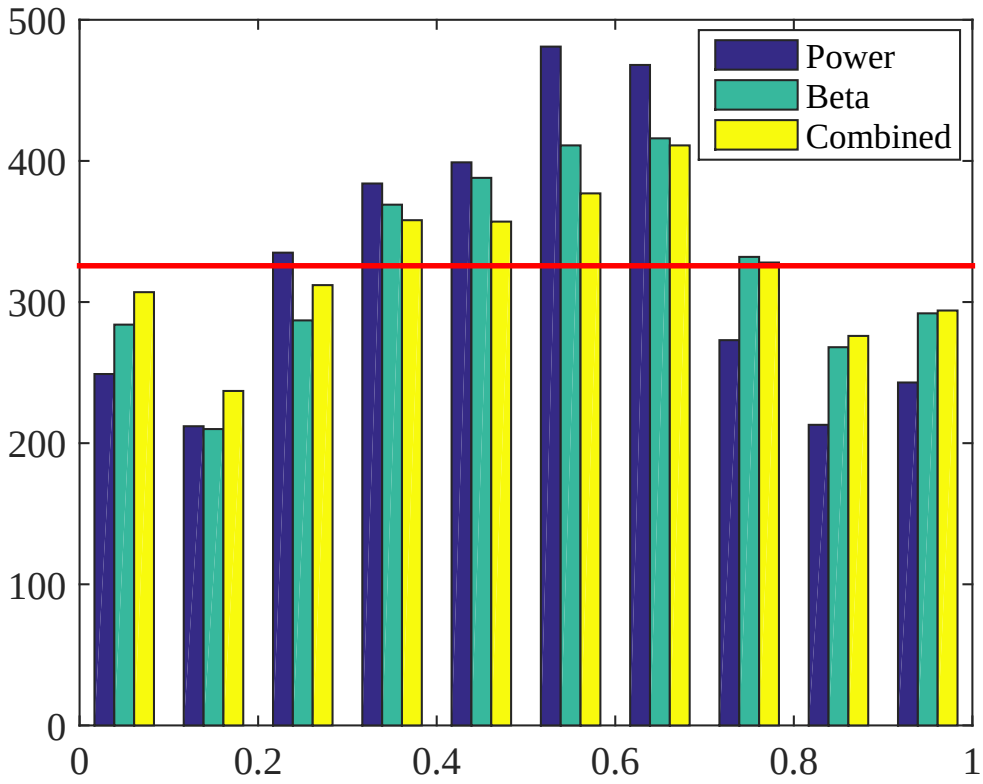

Figure 4.16: PIT histogram for Elizabeth City

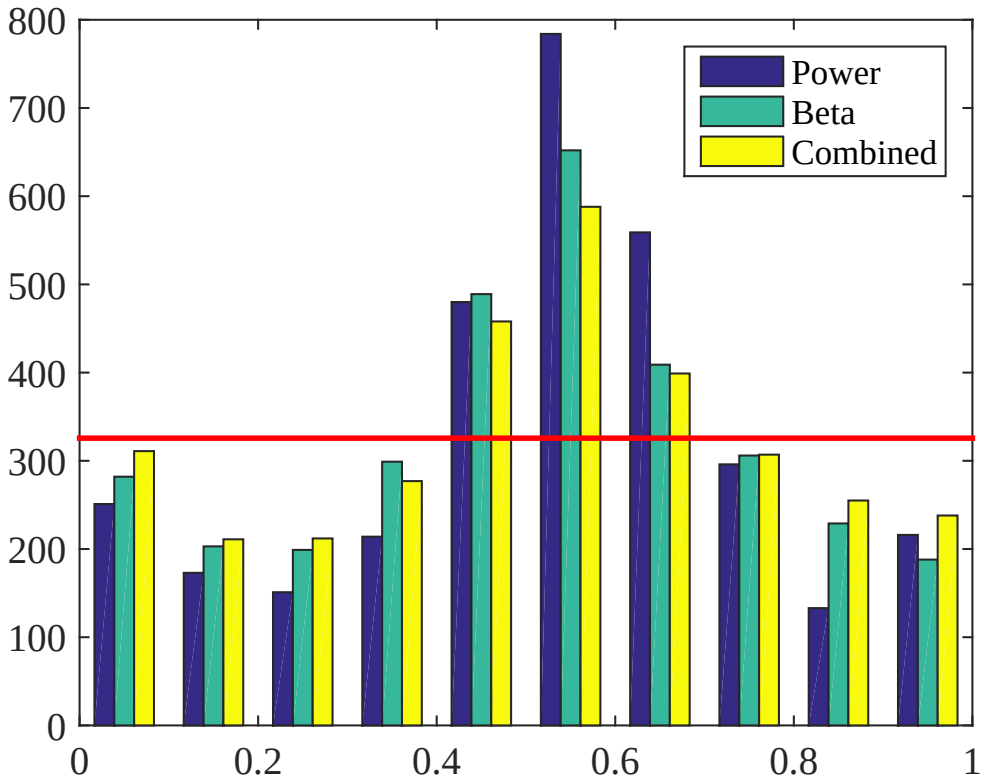

Figure 4.17: PIT histogram for Los Angeles

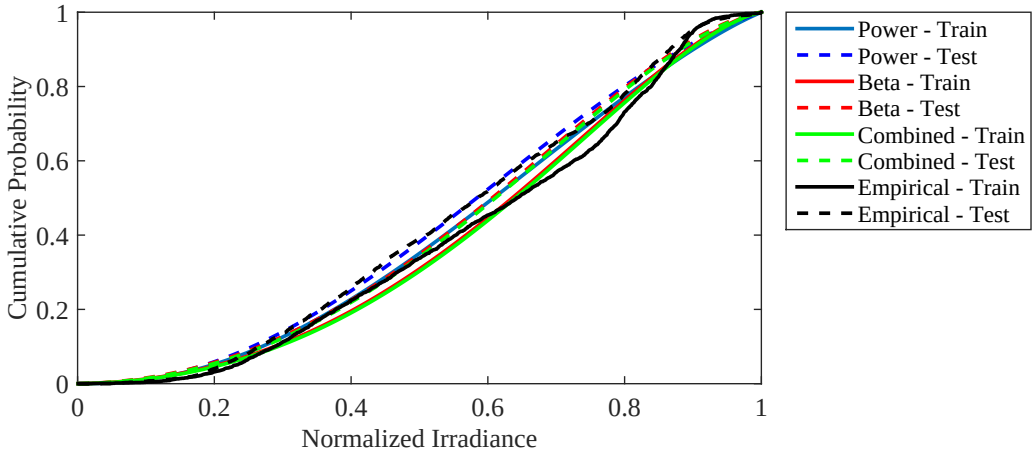

Figure 4.18: Average predictive CDF vs empirical marginal CDF in LaOla

CDF of the observation. We estimate true CDF of observations by empirical CDF of the observations and expected predictive CDF by averaging predictive CDF of each observation. In Fig.4.18, Fig.4.19 and 4.20 average predictive CDF of each method and empirical CDF of the observations for LaOla, Elizabeth City and Los Angeles are compared. Consistent with previous observation, the beta and combined method have better marginal calibration than power method. Also, forecasts for Los Angeles have the worst marginal calibration.

The simulations show strong agreement with test and training results. The results are summarized in Table 4.1. Marginal distribution of the observations is the ideal probabilistic forecast if no more information is given. Therefore, we use the empirical marginal distribution as a baseline for comparison. Marginal calibration is reported by maximum CDF difference between empirical marginal CDF and average predictive CDF. Similarly, Probabilistic calibration is reported by maximum CDF difference between empirical CDF of PIT and CDF of uniform distribution. We normalize dispersion to be the reported variance of the PIT multiplied by 12 so one shows neutral dispersion and less than one means over-dispersion.

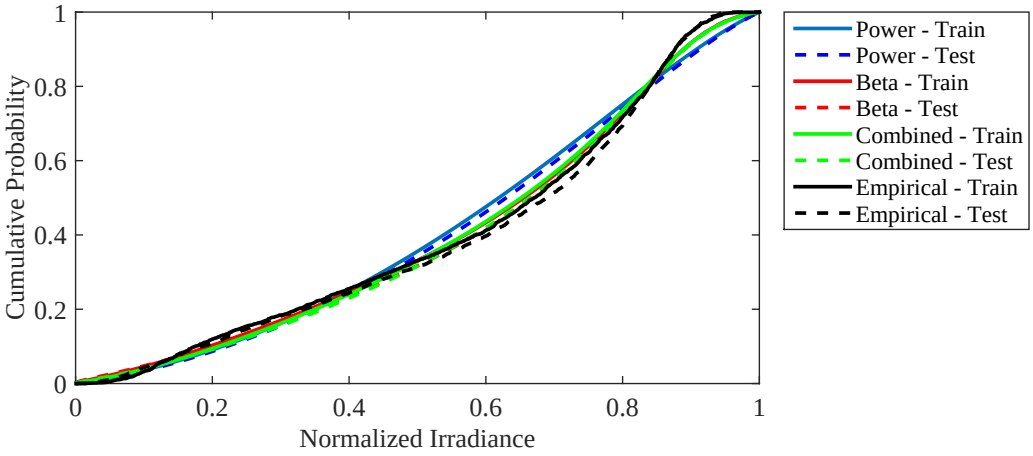

Figure 4.19: Average predictive CDF vs empirical marginal CDF in Elizabeth City

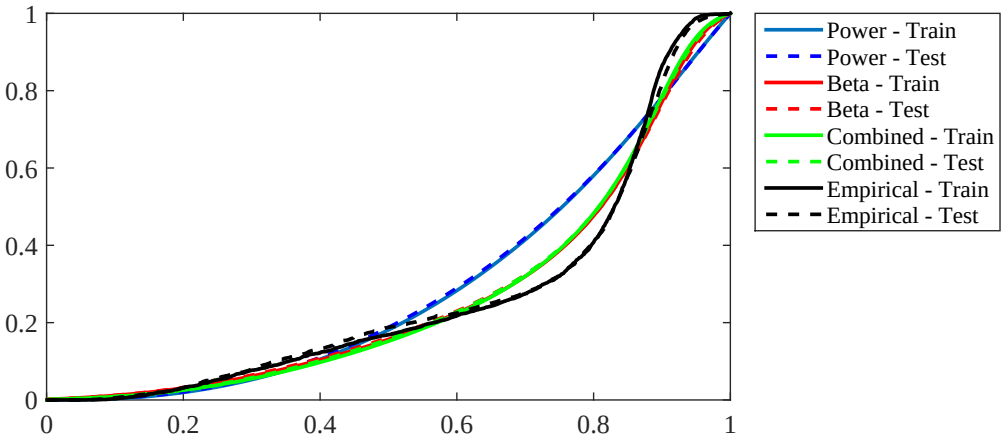

Figure 4.20: Average predictive CDF vs empirical marginal CDF in Los Angeles

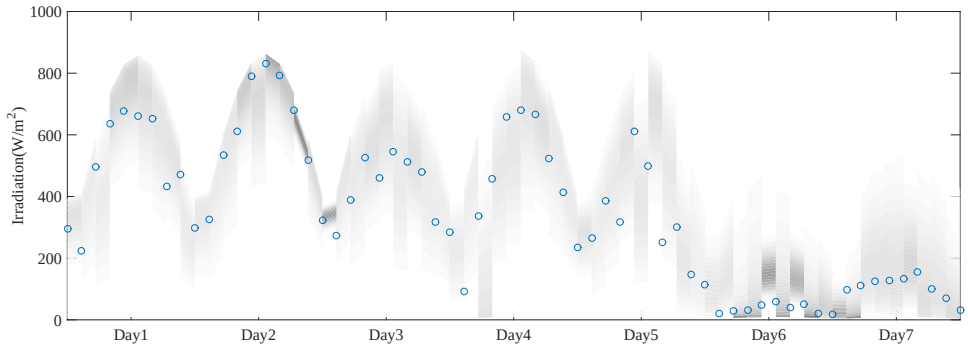

Figure 4.21: An hour ahead forecast for Elizabeth City on March 8 - March 14- 2009

Despite failing in calibration, the forecast for Los Angeles predicts the outcomes better as it has the lowest logarithmic score, smallest CRPS and sharpest interval. On the other hand, the calibrated forecast for LaOla has the worst CRPS and sharpness metric. This is consistent with the observations using point forecasts [32]. Having many partly sunny days in Hawaii, it is harder to predict solar radiation with respect to more stable weather in Los Angeles.

The average logarithmic score for empirical marginal distribution is an estimate of differential entropy and for the ideal forecast is an estimate of conditional differential entropy. By conditioning on past observation, the uncertainty decreases and the lower band for logarithmic score is conditional differential entropy. Hence Los Angeles with -0.59 score intrinsically has less uncertainty with respect to Elizabeth City and LaOla with -0.176 and -0.168 scores. The combined method significantly reduces the logarithmic score for Elizabeth City from -0.168 to -0.669 and for Los Angeles from -0.588 to -0.963. However, there is a lower decrement from -0.176 to -0.338 for LaOla. This can be justified by stronger dependency on past in Los Angeles and Elizabeth City with respect to LaOla.

As an example, an hour ahead predictive probability density for one week after denor-

Table 4.1: Evaluation of different aspects of the probablistic forecast for different sites

Site	Method	Log Score		CRPS		Sharpness		Marginal Cal.		Probabilistic Cal.		Dispersion	
		train	test	train	test	Central 50%	Central 90%	train	test	train	test	train	test
LaOla	Empirical Marginal	-0.1760	-0.1536	0.1279	0.1298	0.4067	0.6933	0	0.0815	0	0.0814	1.00	0.9932
	Power Method	-0.2778	-0.2743	0.1046	0.1046	0.2973	0.6611	0.0692	0.0711	0.0452	0.0527	0.8942	0.8961
	Beta Method	-0.3139	-0.3069	0.1006	0.1014	0.2677	0.5936	0.0496	0.0636	0.0290	0.0441	0.9435	0.9429
	Combined	-0.3378	-0.3215	0.1009	0.1017	0.27	0.61	0.0503	0.0612	0.0532	0.0535	0.9019	0.9014
Los Angeles	Empirical Marginal	-0.5878	-0.5479	0.1096	0.1101	0.2367	0.7100	0	0.0786	0	0.0788	1.000	0.9998
	Power Method	-0.6461	-0.6337	0.0706	0.0710	0.2053	0.4871	0.1920	0.1913	0.1687	0.1678	0.7001	0.7019
	Beta Method	-0.7846	-0.7450	0.0599	0.0607	0.1659	0.3804	0.0908	0.0903	0.0936	0.1023	0.7530	0.7650
	Combined	-0.9630	-0.9341	0.0594	0.0602	0.1447	0.3523	0.0818	0.0770	0.0854	0.0923	0.8202	0.8308
Elizabeth City	Empirical Marginal	-0.1682	-0.1299	0.1409	0.1423	0.4078	0.8044	0	0.0576	0	0.0579	1.000	0.9978
	Power Method	-0.5254	-0.5213	0.0792	0.0798	0.2291	0.5357	0.0904	0.0904	0.0779	0.0795	0.7670	0.7731
	Beta Method	-0.5802	-0.5349	0.0727	0.0740	0.1922	0.4412	0.0301	0.0361	0.0643	0.0637	0.8719	0.8803
	Combined	-0.6691	-0.6365	0.0726	0.0738	0.1792	0.4354	0.0336	0.0410	0.0423	0.0495	0.8977	0.9056

malization of the probabilistic forecasts are shown in Fig. 4.21. The probability density is shown by gray scale color where darker color shows higher probability density. The realized outcomes are shown using ”o” symbols.

4.5 Conclusion

Solar forecasting methods in the literature usually use point forecasts which have a single output value for each prediction, but a single value for prediction of a random variable is not enough for many decision makers and more information about uncertainty of future quantity is required. Probabilistic forecasts can give a more complete stochastic characterization using probability distribution over future quantity (for example predicting a CDF function for a random variable). Recently some papers have discussed probabilistic forecast for solar radiation. However, probabilistic solar forecast is still immature and requires more contributions as combination of different probabilistic forecasts could lead to better a

performance.

Two parametric probabilistic forecast methods based on beta and power distribution are presented to predict solar irradiation and their performance are evaluated. A combining procedure using the beta transformed linear opinion pool is utilized to improve the initial forecasts. Our simulations on three different sites show that these methods reduce the uncertainty by providing lower logarithmic score with respect to baseline forecast which uses the marginal distribution. Calibration metrics shows that these methods despite the simple structure accurately describe the stochastic characterization of solar irradiation. We obtained this result by using simple zenith angle normalization and tap delay filters to find conditional mean and conditional variance. Many other point forecast methods which are minimizing RMSE or MSE can be used for estimation of conditional mean and variance to obtain a probabilistic forecast based on the approach proposed here.

Chapter 5

An Optional Insurance Policy for Reliability of the Electric Grid

5.1 Introduction

In the last several decades there have been many major blackouts. Each particular blackout can be explained with an initial failure followed by a chain of causal events which ultimately leaded to a major system collapse [93]. For example, the 2011 Southwest blackout initiated by the loss of a single 500kV transmission line. The power redistributions, voltage deviations and resulting overloads created a domino effect tripping transformers, transmission lines, and generation units in a very short time. The whole power grid collapsed within eleven minutes after initial failure. The blackout left more than 2.5 million consumer without power and it took about 12 hours to reconnect consumers[94]. Similarly, northeast blackout in 2003 that affected 55 million people can be explained by an initial cause and a chain of cascading events [93].

When a consumer is connected to the electric grid it is the Utility's responsibility to provide continuous power available to the consumers. The costumers may suffer losses due to interruption of the electricity. The annual cost of power interruption to the electricity consumers in the US is estimated to be around $79 billion and due to uncertainty in analysis it could be as high as $133 billion or as low as $22 billion[47].

While each black out occurrence could be explained by an initial failure and a chain of causal events, it is impossible to determine the initial failure beforehand. Hence the initial failure usually is modeled as a random event. Of course, 100% reliability cannot be achieved as human errors or acts of nature like hurricanes, earthquakes and other weather related events cannot be eliminated. However, provisions to provide adequate levels of reliability is possible at a certain cost [95].

Generally, provisions of higher levels of reliability costs more for the Utility; on the other hand, it reduces aggregate losses by the costumers. The reliability level is optimum if total cost of provision and aggregate costumers' losses are minimized[96].

For efficient allocation of reliability two aspects must be investigated. The first aspect is the cost of providing reliability. Generally speaking, having more reserves and upgrading infrastructure (for reliability purpose) will result in higher reliability levels and the cost associated with them could be clearly determined. The second aspect is reliability worth or how much the society is willing to pay for the reliability. Since reliability services are currently combined with electric energy as a private good, its value can not be determined directly using market transactions[75–77, 97]. The methods for estimation of outage costs can be categorized by three general approaches; macroeconomic, survey based and market based approaches [98].

The macroeconomic approaches consider dependency of macroeconomic indicators

such as the gross domestic product to the electricity and make conclusions based on value added to society and the amount electricity usage. One advantage of this approach is that the conclusions reached by these studies are not limited to specific blackout circumstances and the results are applicable to more general situations. However, different individual blackout factors are not generally taken into account and direct damages are neglected in these studies [97].

The survey based approaches rely on surveys from interviewed consumers to evaluate outage costs by analyzing responses to questions. In these surveys, hypothetical power outage scenarios are used to collect the necessary data. The consumers are asked to either directly estimate outage cost or to select between different scenarios and interruption costs are later calculated from their choices[99, 100]. The advantage of survey-based studies is in the freedom of the hypothetical scenarios. In these survey-based studies, it is possible to take different outage conditions and consumer damage factors into consideration without being restricted to available alternatives or actual power outages. However, the results of these studies suffer since they do not consider actual choices and in real situation people may show different choices.

Market-based approaches estimate the economical value of the power not delivered to the consumer (value of lost load or VoLL) by analyzing consumer preferences on the basis of observations of real consumer choices made in the past regarding service reliability. The observations can be either before or after of power outages. According to Sullivan and Keane [101] results of market-based studies has high validity since they rely on actual economic choices. However, the degree of freedom in market-based studies is limited because results are derived from choices based on available alternatives.

There are many factors that influence outage costs. For example duration of outages,

time of outage, prior notice. Different individuals also have different cost factors as the discomfort level is different based on age, gender and income [75–77, 97].

Currently, electricity and reliability service are bundled together and enforce costumers to pay for the bundled price. It is clear that there is a trade-off between immediate operating cost and the long term goal of reliability as electricity and reliability are bundled together. Since the Utility hardly receives any signal about consumers' preferences on reliability, the immediate cost of operation is the winner and reliability levels maintained (hopefully) at minimum mandated standards (for example N-1 secure constraint). So the reliability upgrades are postponed to the future and the reliability of the system constantly decreases as infrastructure deteriorates and demand increases over time. The degradation of the system continues until a big blackout happens in the over stressed system and the deficiencies of the power system are highlighted requiring immediate substantial upgrades. The cycles of degradation and substantial upgrades create a double paradox effects since the societies who are used to high levels of reliability experience more losses due outages as they are less prepared for it and less concerned about it [16].

In this chapter, we argue that reliability of electric grid is a public good and we use an insurance framework to implement a benefit taxation mechanism that provides a framework to achieve optimal reliability levels.

5.2 Terms and Definitions

— *Rival:* a good is rival if its consumption by one consumer prevents simultaneous consumption by other consumers.

— *Excludable:* a good or service is excludable if it is possible to prevent consumers

who have not paid for it from having access to it.

— *Public Good* is a good that is both non-excludable and non-rival in that individuals cannot be effectively excluded from use and where use by one individual does not reduce availability to others.

— *Private Goods* are goods which are both excludable and fully rival.

— *Congestible Public Good* are public goods where, after a point, the enjoyment received by the consumer is diminished by crowding or congestion. For example parks and highways.

— *Willingness to Pay (WTP)* is the maximum price at or below which a consumer will definitely buy one unit of the product.

— *Pareto Efficiency* is a state of allocation of resources in which it is impossible to make any one individual better off without making at least one individual worse off.

— *Benefit Taxation* is a form of taxation system conceived by Erik Lindahl in which individuals pay for public goods according to their marginal benefits.

— *Strict Liability* is a form of liability that makes a person legally responsible for the damage and loss caused by his/her acts and omissions regardless of culpability. Under strict liability, there is no requirement to prove fault, negligence or intention.

— *Negligence Liability* is a form of liability that makes a person legally responsible if he/she has failed to use reasonable care, resulting in damage or injury to another.

— *N-1 secure* is situation of the system in which service is still provided under any single contingency event.

— *N-2 secure* is situation of the system in which service is still provided under any combination of two single contingency events.

In the current electric system, once a costumer is connected to the electrical grid, it is not possible to disconnect her/him from the grid and system operators can only disconnect loads on emergencies at substation levels rather than customer level. The system operators are also mandated by law to provide continuous power to all the customers. (In the future, if system operator can communicate and/or disconnect at the customer level using smart meters then the reliability service also becomes excludable i.e. a private good. However, this idea is currently impractical.)

In the normal conditions, the usage of one consumer has infinitesimal effect on the other consumers, however, in emergency conditions (when insufficient generation is available due to a contingency event or extremely high demand) the consumption of one consumer has adverse effects on the usages of others. Therefore, reliability of the power system is a congestible public good.

5.3 Limitations of Bundled Pricing

Electricity service is provided as a package to the consumers including all generation, transmission, distribution, ancillary and other services necessary to deliver and measure useful electric energy and power to consumers. The total cost of providing the package is usually distributed between the consumers based on their consumption. Therefore, all costumers pay for a bundled price. While the power and energy in the package are private goods, the services like voltage stability, frequency stability and reliability are considered as public goods. Therefore the bundled pricing has several limitations.

Using a bundled pricing policy, the reliability will not increase, no matter how much the utility receive money from customers. This fact has been shown in [102, 103]. Kirschen

and Strbac showed that unless the governing rules of the system usage is changed, additional transmission capacity would be available to all users of the transmission network. Hence they demonstrated using a simple example that these investments do not decrease blackout probability [102]. Reppen also states that under N-1 criteria the reliability of the system constantly degrades without showing itself in the criteria [103]. This means that over the time the existing reliability margins in the system are utilized which increases the probability of black out; however, no investment will be done since system is still N-1 secure.

Bundled pricing policy is not optimal for either customer or for the Utility. Considering the benefits for the Utility when costs are raised. Those customers who have lower WTP for reliability decrease their electricity consumption which reduces expenditure on both electricity and reliability and those who have higher WTP for reliability may still be paying less than what they are willing to pay. If those consumers who use more electric energy are also demanding more reliability, we can make a connection between energy usage and reliability demand as a possible proxy to estimate willingness to pay so the bundled price becomes socially optimum for both reliability and electricity. However, the demand for electricity and reliability are not always connected together. Some firms may use small amount of energy while they require high reliability; on the other hand, some firms may use large amount of energy but are less concerned about reliability. For example, for agriculture crops a large amount of energy is used for irrigation purposes but the crops can easily tolerate energy outages by rescheduled watering. Similarly, water supply and sewage sector demands for high energy quantities but is less vulnerable to potential outages. On the other hand, construction and wholesale sector need high reliability [16].

Bundled pricing is not optimal for customers, as well. When a consumer is connected

to the power grid it is the responsibility of the Utility to provide continuous power available to the consumers. The costumers may suffer losses due interruption of the electricity. Consumers may claim compensation for the damages and losses and the Utility is often mandated (for example Rule 16 in Hawaiian Electric Company) to compensate the consumers losses if it is under their control. However most of the time, the main cause of the blackout is an initial failure which is usually out of control of the Utility company and hence the customer will not get any compensation. Many court cases are reviewed in [104]. In such a framework there is no incentive for the Utilities to invest against events outside their control for example investing in underground transmission lines rather than overhead transmission lines in areas with higher storm probability. Strict liability for the Utility may also seem a solution for this problem. However, in that case the Utility will over invest on reliability to minimize its own cost and force the costumers to pay for it, which leads to a non-efficient solution[105].

Customers with higher WTP may buy their own back up generation and pay for the capital and maintenance of it to get to reliability they need. However considering skill and specialization factor of the Utility and economies of scale this solution is not efficient. Another way for costumers with higher WTP is to buy an insurance from independent insurers and transfer risk of business interruption due to power outages, but in such cases insurer cannot fight against the outage causes. The insurer only estimates the distribution of losses and based on that computes and collects the premiums and compensate the losses if it happens. So both of aforementioned solutions are not optimal. We propose an "optional insurance policy" in the following section.

5.4 Decoupling Reliability and Electricity

Let us assume the Utility has an account to insure against outages by collecting optional premiums from the consumers. If no one wants to pay the premium it means that current reliability situation is acceptable and extra care and investment is not required. On the other hand if some consumers are willing to pay the premiums, this means that current reliability situation needs to be improved and the Utility company has incentive to spend some of accumulated premiums for reliability upgrade in order to reduce outage probability and keep the rest for compensation of potential future outages.

Let us consider a simple example, assume we have two areas A and B which are connected using three transmission lines with each having 300MW capacity. There is an efficient generator at area A with marginal cost equal to $20 per MWh and another generator at area B with marginal cost equal to $50 per MWh. A load equal to 1200MW is located at area B and both generators have enough capacity to serve entire load. The cheapest operation is to assign full transmission capacity (3*300MW=900MW) to the cheap generator and the remaining 300MW to the expensive generator (Fig. 5.1). This scenario costs $33000 per hour but if for any reason one of the transmission lines get tripped both other lines are overloaded and get tripped and the system collapses. To provide more reliability, 600MW is assigned to the cheap generator and 600MW to the expensive generator (Fig. 5.2). In this case if one of the lines fails, its power is transferred to the other two lines and system continue to work but this scenario costs $9000 per hour more. We use the N-1 secure system as the baseline.

Let assume that capital invested to increase capacity of all three lines from 300MW to 400MW. In this case, system operators consider new N-1 scenario and assign 800MW to

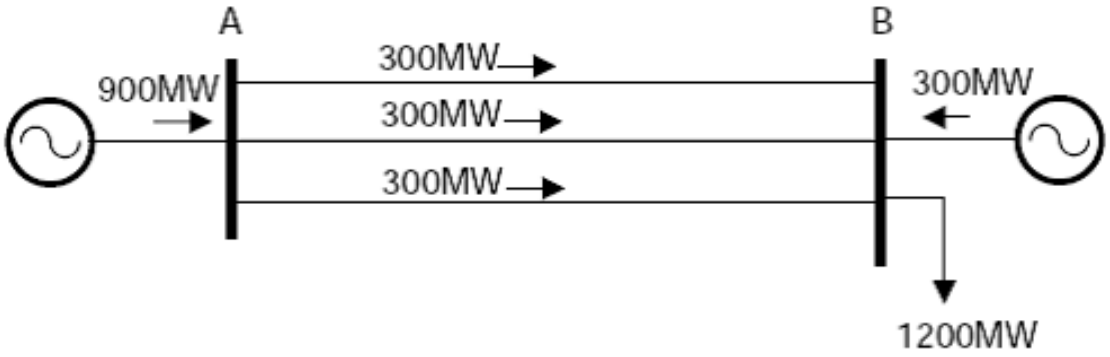

Figure 5.1: Cheapest but non-secure scenario.

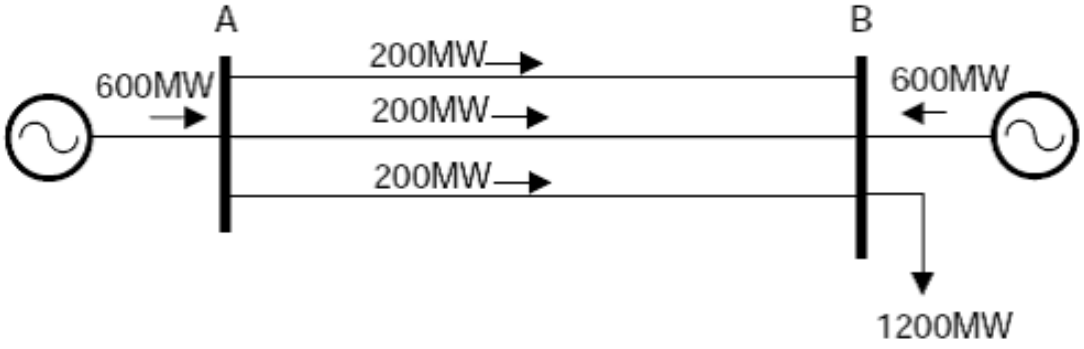

Figure 5.2: N-1 secure operation.

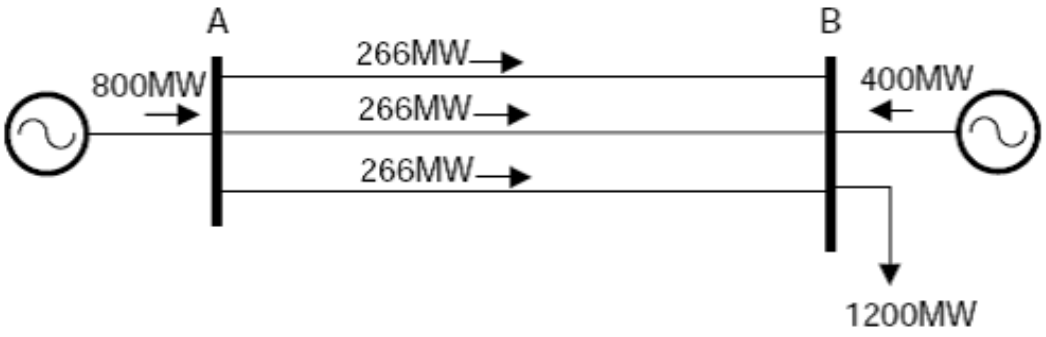

Figure 5.3: When reliability is mandated by N-1 criteria, increasing transmission capacity, decreases the operation costs but reliability does not improve.

the cheap generator and 400MW to the expensive one as this scenario is the cheapest way which is consistent with the standards. It is clear that operation cost is decreased but the reliability stays the same as before.

If instead of increasing capacity of existing lines a new transmission line with 300MW capacity is created. In this case system operator assigns 900MW to cheap generator and 300MW to the expensive one as the cheapest way to operate the system in compliance with the standards. In this case, addition of new line decreases operating cost and N-1 security stays the same, however, the reliability is decreased since probability of outage of more than one line among four lines is more than probability of outage of more than one line among three lines. We note that if new line is exclusively used for the reliability purpose (i.e. four lines each carries 150MW as shown in Fig.5.5) system become N-2 secure and reliability substantially increases.

If system (without any improvement) runs with N-1 secure strategy and p_0 is probability of a line outage during a 30 days billing time. The system will collapse if two or more line get tripped so it happens with $3p_0^2 - 2p_0^3$ but for N-2 secure scenario the blackout

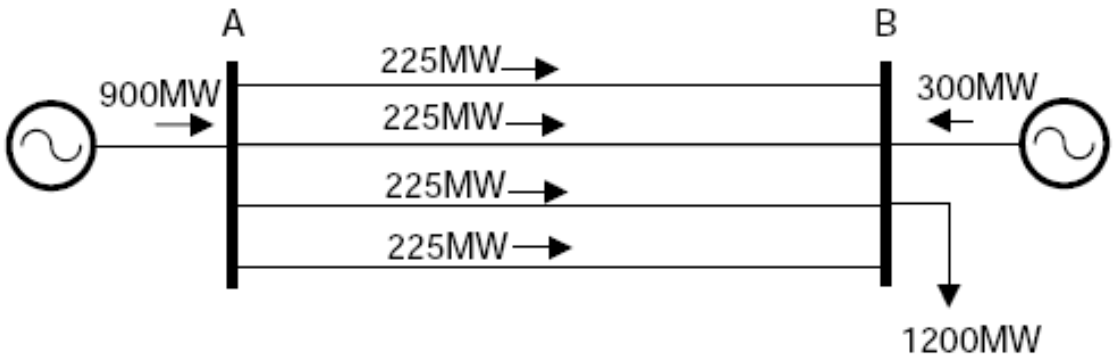

Figure 5.4: When reliability is mandated by N-1 criteria, adding a new transmission line will decrease the operating cost; however, the reliability percentage become worse as more lines will be subject to a possible failure.

probability is only p_0^3. So upgrading to a N-2 secure system decreases the risk by a factor of $p_0/(3-2p_0)$. Hence upgrading system to N-2 secure system is justified if the total collected premium times $1 - p_0/(3 - 2p_0)$ is more than the extra cost for operating system with N-2 secure scenario ($6480000 here). If the total premiums is not enough to immediately run the system with N-2 scenario, the Utility keeps the money in an account to accumulate for potential loss compensation or future reliability upgrade. If a new line is built only for reliability improvement, the risk of blackout decreases by a factor of $(3p_0 - 4p_0^2)/(3 - 2p_0)$. Therefore if the accumulated money times $1 - (3p_0 - 4p_0^2)/(3 - 2p_0)$ is more than construction cost of a new transmission line, the Utility will build a new line and use it only for the reliability purpose (i.e. runs the system with four lines with N-2 secure as shown in Fig.5.5).

Compensations are computed by amount of premium that is paid by each customer divided by outage probability in the base case. Risk neutral decision makers compute their

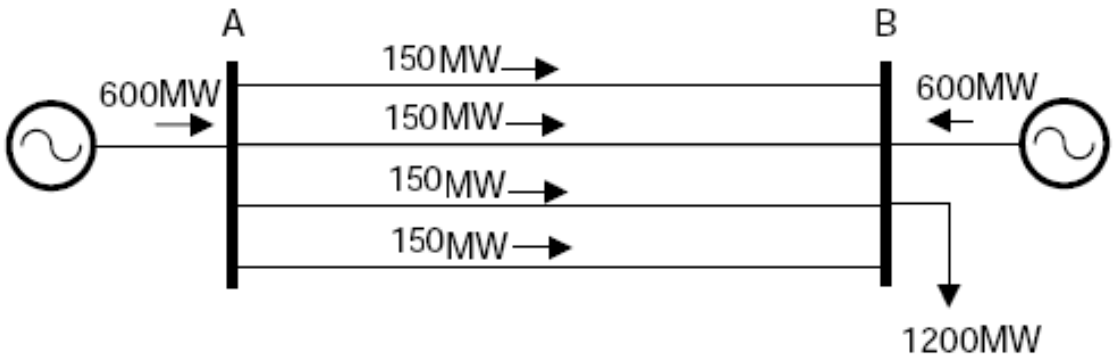

Figure 5.5: If a transmision line is built only for reliability purpose.

premium based on their expected losses and risk adverse decision makers willing to pay even more. So the total collected premiums is a measure of expected aggregate losses. Based on results of extensive surveys, people do not consider reliability as a new service and value it based on expected (direct and indirect) losses [99, 100]. Therefore the total collected premiums is aggregate willingness to pay for reliability of power system. Those who are not willing to pay premiums will not compensated when outage losses occur.

Reliability worth also depends on time and different consumers evaluate it differently based on time. For example, some business may value it more during working days others may value it more during weekends and holidays. In the optional insurance policy, the customers can distribute the premiums according to their desires and therefore we can address variation of the optimal reliability accordingly.

Coming up with an optional insurance policy for reliability may seem similar to demand response problems since in both cases flexible costumers will pay less than others. But optional insurance policy is different in many aspects. In demand response programs, consumers respond to a price signal and decide whether to consume power or not. The

participation in demand response programs are usually low as people have more important things to decide and so they usually disregard the price signal unless an automatic decision making system is available. But in the insurance framework, people only make a long term decision once and forward their decisions to the Utility. The aggregate premiums show the willingness to pay for the reliability. The Utility could spend part of this money for improving reliability and the remaining part will be accumulated for reimbursing future outage losses. In fact by considering this insurance framework, the customer actively participates in the Utility investments, rather than active participation in demand management.

5.5 Conlusion

While the active and reactive power which is delivered by the power grid is a private good because it is both excludable and rival, other characteristics of this private good such as voltage and frequency stability and system reliability are considered as public goods. So a method is required to make the social optimal decision about quantity of these services. The commonly used rule of thumb is that outages should happen less often than 1 day in 10 years. However this rule of thumb may be different than the social optimal decision; hence, researchers defined an economic criteria called value of lost load or VOLL which is used to estimate the marginal social cost of unserved loads. VOLL can be used by decision makers to decide on optimal quantity of electricity reliability however with current pricing method which bundles the price of electricity and other services and a unique price is for this mixed good is not efficient as the VOLL varies for different firms.

The proposed insurance insurance policy is a way to decouple electricity and reliability and a proxy for implementation of benefit taxation which can achieve both optimal quantity

and optimal price. The premium of the insurance is optional but depends on two factors, outage probability and losses by the costumers if they experience an outage. The probability of the outage is computed by the Utility and need to be approved by Public Utilities Commission. Each costumer pay the premium he or she will to pay but the compensation is based on amount of premium divide by the probability of an outage. Therefore, a risk neutral decision maker will to pay a premium equal his/her losses times the outage probability.

Chapter 6

Energy Efficient Scheduling Algorithms for Pumping Water in Radial Networks

6.1 Introduction

Electricity cost of pumping water is a large part of the total operation costs of water supply networks [17]. In recent years research has been conducted in [18–22] to model the water network and to reduce costs. Different optimization algorithms (linear Programming [22, 23], quadratic programming, dynamic programming, genetic algorithm[24], ant colony[25], heuristics [26]), various hydraulic modeling (nonlinear hydraulic[27], simplified hydraulic[28, 29], mass transfer[30]), and different decision variables (pump operating time, tank levels) [31] have been explored. In most of the past research, users do not have any active participation and demands are either modeled as a deterministic model or a stochastic process that is predicted using a time series model. In this chapter we consider active participation for the users of a radial network and we show that suitable collaboration

of users can reduce the required energy to pump the desired amount of water for all users. We use a nonlinear hydraulic model for the pump (both fixed speed and variable speed) and pipes and we find the most energy efficient schedule for two users using nonlinear optimization. The solution for the two user case, is used to find an approximate solution for more than two users.

6.2 Modeling

6.2.1 Model for the Pumps

The amount of power that a pump uses depends both on pump parameters such as efficiency and the Q-H curve associated with the pump, as well as the operating point of the pump (that, in turn, depends on the fluid demand, network characteristics, and supply schedule). We begin with the characterization of the pump parameters. A Q-H curve for a pump relates the amount of flow (Q) through the pump to the head, or the pressure difference between the input and the output of the pump. A common approximation is to assume the Q-H curve to be considered is a quadratic function of Q and H for a specific rotation speed:

$$H(Q) = H_0 - K_D Q^2, \tag{6.1}$$

where H_0 is maximum head which the pump can generate and K_D is a parameter that determines the trade off between head and flow rate and is obtained through curve fitting.

Given this curve, the power usage of the pump is then given by

$$P = \frac{\rho g Q H(Q)}{\eta} \tag{6.2}$$

where ρ is the specific gravity of the fluid being pumped, η is the efficiency of the pump and g is the gravitational constant. The efficiency of the pump is not a constant; rather it depends on the operating point of the pump. Thus, a pump is associated with both a Q-H curve and additionally the efficiency or equivalently braking horse power (BHP) curves. In a BHP curve, the total power consumption of the pump as a function of discharge flow is measured and plotted (see, e.g. Fig. 6.1). The BHP curve is usually modeled as a cubic function of the flow [106]:

$$BHP(Q) = P_0 + P_1 Q + P_2 Q^2 + P_3 Q^3 \tag{6.3}$$

where P_0 is the shut off power (power consumption when the outlet of the pump is completely shut off), which measures constant losses such as disk friction and intercept for other hydraulic losses such as recirculation [107]. P_1, P_2 and P_3 are coefficients that account for the useful power of the pump and changes in the hydraulic losses (flow friction, shock, leakage and recirculation) as flow increases. We note that P_3 is negative which results in the power consumption decreasing when the pump is operated at high flows but low heads.

The Q-H and BHP curves of the pumps are obtained through experiments for various values of the rotation speed of the pump. If the electric motor which drives the pump rotates at a nominal speed, the Q-H curve is constant. However, if the motor is equipped with variable speed drive (VSD), both braking horse power and head-flow curves change according to rotation speed (see Fig. 6.2). If equation (6.1) and (6.3) give the pump performance curves at nominal speed, the performance curves at new speed can be computed

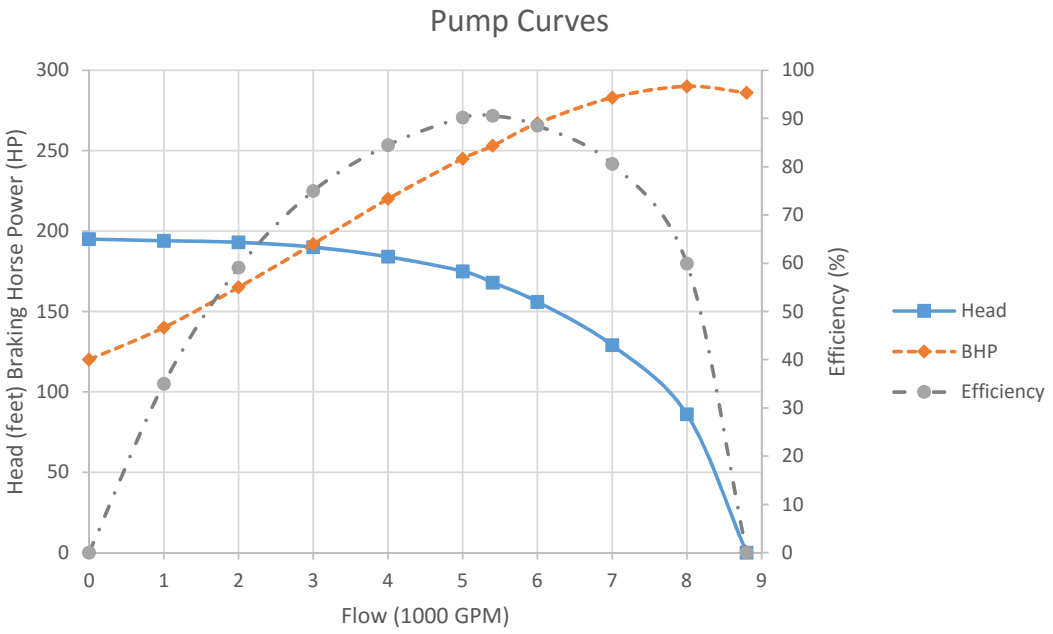

Figure 6.1: Performance curves of a 12" pump at 1180 rpm based on affinity laws [107, 108].

$$H(n,Q) = H_0(n/n_0)^2 - K_D Q^2 \tag{6.4}$$

$$BHP(n,Q) = P_0(n/n_0)^3 + P_1(n/n_0)^2 Q + P_2(n/n_0)Q^2 + P_3 Q^3 \tag{6.5}$$

where n is rotation speed and n_0 is nominal rotation speed (usually is maximum rotation speed).

6.2.2 Model for the pipe networks

Modeling the power loss due to friction as the fluid is pumped through a pipe is a complicated function of the pipe parameters and the operating conditions. This power loss is usually modeled through the head loss. The Hazen-William equation empirically relates

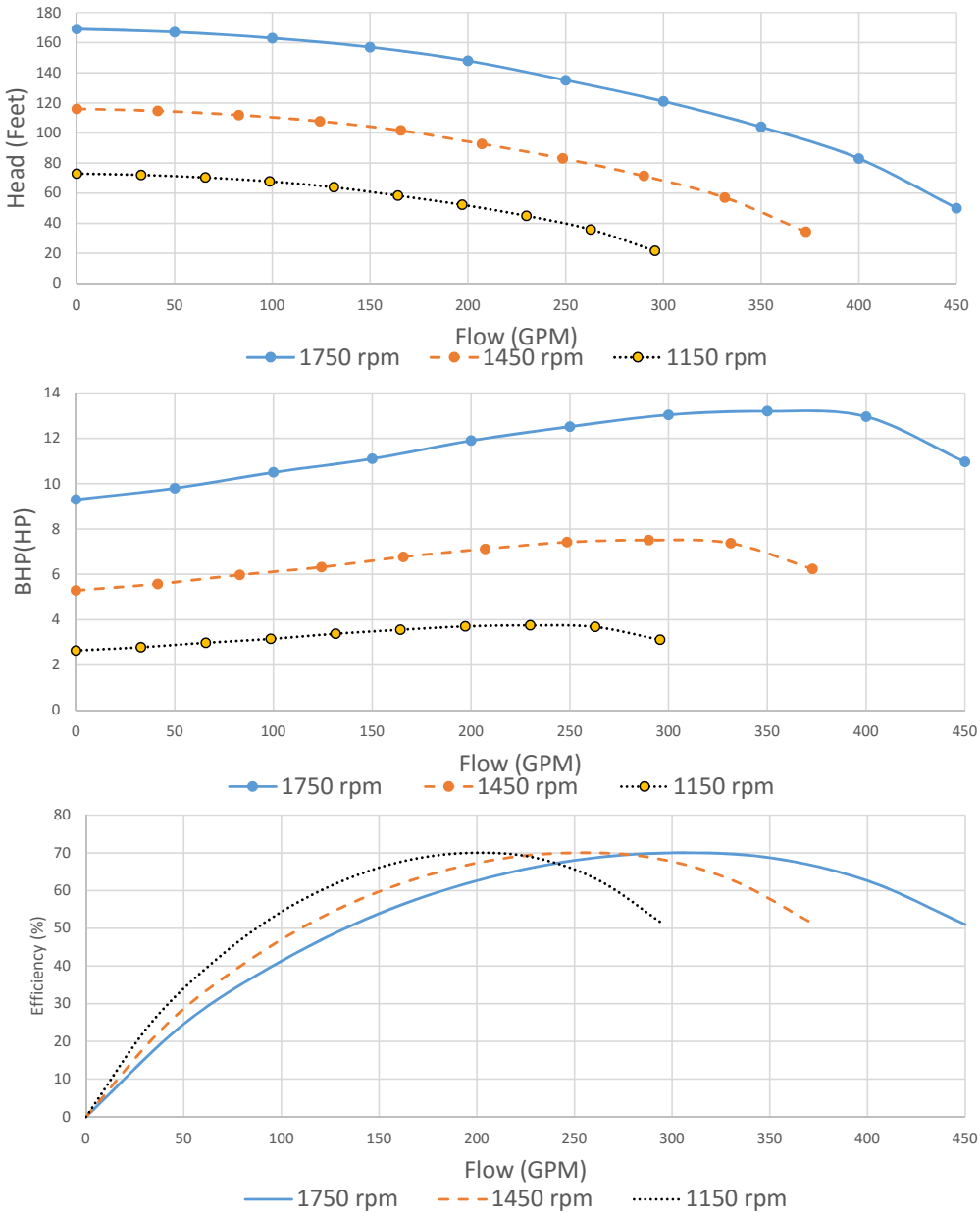

Figure 6.2: Performance curves of a 3” Pump at different rotation speed

the flow of water in the pipe and the head loss. While it is a simple formulation, it is known to be rather inaccurate. Due to its limitations, however we use the more general (and usually preferred [109]) Darcy-Weisbach equations in this paper. According to this equation, the amount of head loss is proportional to square of the flow in the pipe.

$$H_{loss}(Q) = KQ^2, \tag{6.6}$$

where K is a coefficient which is determined by parameters such as the length, diameter and shape of the pipe and the Reynolds number of the fluid being pumped [1]. It is interesting to note that a network of pipes can be modeled as a resistive electric circuit, in which the role of current is played by the flow and that of voltage by the head. However, while analogies of Kirchoff current and voltage laws hold, the linear Ohm's law in the electric circuit needs to be replaced by the non-linear relation in (6.6). We note that with this analogy, power is given by head times flow.

6.3 Problem Statement

Consider a radial network of N users. The network can be described by a linear graph with $N + 1$ nodes. Let the nodes be numbered 0, 1, 2, $\cdots$ from one end with the node 0 denoting the pump. Let node i correspond to user i who requires a certain volume V_i of water over a time period T. This volume can correspond, e.g., to water needs over a day that begins at time T. Any water supplied over the time $[0, T]$ can be stored in a local tank and used over the day.

1. We note that dependency of the coefficient to the Reynold number makes it dependent to flow, so several recalculations are required to tune the final solution

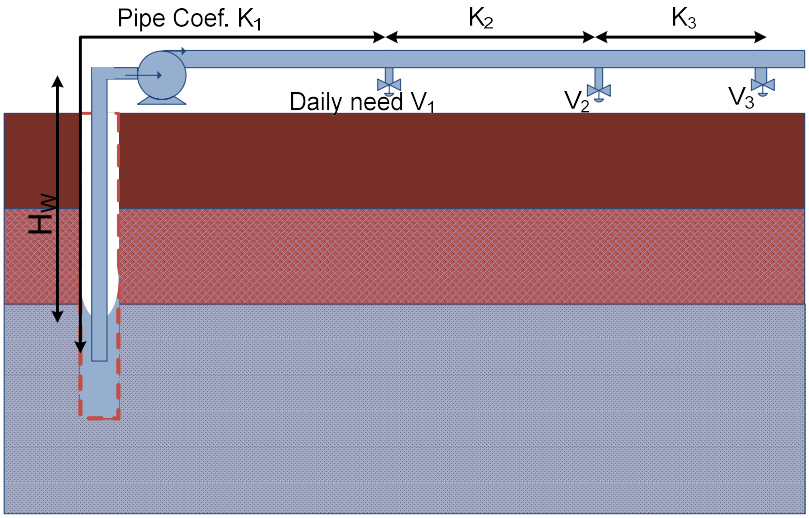

Figure 6.3: Scheduling problem

We assume that the pump supplies water from a well with water level at an elevation H_w under the ground. The pump performance curves are given by equations (6.1) and (6.3).

Given the non-linear dependence of head loss in a pipe and power usage of the pumps on the flow, it is clear that scheduling of water for various users will have a significant effect on the total energy consumption. The optimization problem that we seek to solve is to design the water supply schedule to minimize the total energy consumption at the pump while meeting all the user demands in terms of the volume of water needed by time T.

6.3.1 Model for required energy to supply a radial network with a fixed speed pump

If the pumps is not equipped by variable speed drive it turns at relatively constant speed and operating point is on the Q-H curve. This may create more head than the required head for a specific pipe and flow. The extra head wastes the energy in the pipes and can cause

higher pressure at the outlets. The total energy usage is given by:

$$E_{total} = \int_0^{T_1} P_0 + P_1 \sum_{i=1}^{N} q_i(t) + P_2 (\sum_{i=1}^{N} q_i(t))^2 + P_3 (\sum_{i=1}^{N} q_i(t))^3 dt \tag{6.7}$$

where $q_i(t)$ is the flow for the i-th user at time t and T_1 is total time that pump is on in the period T. The schedule should satisfy demand of all users:

$$\int_0^{T_1} q_i(t)dt = V_i \quad 1 \leq i \leq N \tag{6.8}$$

Also, the head which is generated by the pump should be greater than head required by the pipe networks at all times:

$$H_0 - K_D (\sum_{i=1}^{N} q_i(t))^2 \geq H_w + \sum_{j=1}^{N} K_j (\sum_{i=j}^{N} q_i(t))^2 \quad 0 \leq t \leq T_1 \tag{6.9}$$

We also assume that users only consume water and cannot return water to the pipe so demand is nonnegative for all users at all time.

$$q_i(t) \geq 0 \quad 1 \leq i \leq N \quad 0 \leq t \leq T \tag{6.10}$$

We therefore have a constrained nonlinear optimization and the discharge flow at different outlets are the decision variables. The pump turns off if there is no discharge flow and turns on if one of the decision variables become positive.

6.3.2 Model for required energy to supply a radial network with a pump with VSD

If the pump is equipped by a variable speed drive (VSD), a controller adjusts the rotation speed of the pump to avoid creating excess pressure on the pipes. So we have a

different optimization since we have both rotation speed ($n(t)$) and discharge flows as decision variables.

$$\min_{q_i(t),n(t)} \quad \int_0^{T_1} P_0(n(t)/n_0)^3 + P_1(n(t)/n_0)^2 \sum_{i=1}^{N} q_i(t) + (n(t)/n_0) P_2(\sum_{i=1}^{N} q_i(t))^2 + P_3(\sum_{i=1}^{N} q_i(t))^3 dt \tag{6.11}$$

$$\text{s.t.} \quad \int_0^{T_1} q_i(t)dt = V_i \quad 1 \leq i \leq N \tag{6.12}$$

$$H_0(n(t)/n_0)^2 - K_D(\sum_{i=1}^{N} q_i(t))^2 \geq H_w + \sum_{j=1}^{N} K_j(\sum_{i=j}^{N} q_i(t))^2 \quad 0 \leq t \leq T_1 \tag{6.13}$$

$$q_i(t) \geq 0 \quad 1 \leq i \leq N \quad 0 \leq t \leq T \tag{6.14}$$

6.4 Solution formulation

6.4.1 Fixed speed pump

Optimal schedule for one user: The optimal solution is to fill the capacity of the pump so that there is no excess pressure created and the task is done in minimum time.

$$q_1(t) = \begin{cases} q_{max} = \sqrt{\frac{H_0 - H_w}{K_D + K_1}} & 0 \leq t \leq V_1/q_{max} \\ 0 & \text{otherwise} \end{cases} \tag{6.15}$$

If q_{max} is smaller than the best efficiency flow, the solution is intuitive since in addition to using the pump for less time, we also approach the best efficiency point. When q_{max} is larger than the best efficiency flow, the pump works at a less efficient point but turns on for less time and this results in lower pressure loss at discharge outlet which overcomes the decrease in efficiency.

Optimal schedule for two users: In a radial network the users jointly own the pump and pipe and they should determine how to share their usage. Each sharing strategy has its own benefit and cost. For example, if users use the pipe and pump together power loss in second pipe decreases and pump may operate at more efficient point, but power loss due to extra pressure at the first discharge outlet is unavoidable. In this section we find a partial sharing strategy that solves the trade off between sharing or no-sharing to find the schedule that minimizes energy usage of the pump. In this partial sharing strategy first user partially opens its valve and the second user completely opens its valve until the first user needs are satisfied, then the second user uses all the flow until the second user needs are completely satisfied. For extreme points the partial sharing corresponds to a full share or no share.

$$q_1(t) = \begin{cases} Q_p \frac{V_1}{V_1+\alpha V_2} & 0 \leq t \leq \frac{V_1+\alpha V_2}{Q_p} \\ 0 & \text{otherwise} \end{cases} \tag{6.16}$$

$$q_2(t) = \begin{cases} Q_p \frac{\alpha V_2}{V_1+\alpha V_2} & 0 \leq t \leq \frac{V_1+\alpha V_2}{Q_p} \\ Q_2 & \frac{V_1+\alpha V_2}{Q_p} < t \leq \frac{V_1+\alpha V_2}{Q_p} + \frac{V_2(1-\alpha)}{Q_2} \\ 0 & \text{otherwise} \end{cases} \tag{6.17}$$

where Q_p is flow of the pump when both users use the water, α is fraction of V_2 which is shared with first user and Q_2 is maximum flow if only the second user uses water. Then, the optimization problem becomes:

$$\min_{Q_p,\alpha} \quad (P_0 + P_1 Q_p + P_2 Q_p^2 + P_3 Q_p^3)\frac{(V_1+\alpha V_2)}{Q_p} + (P_0 + P_1 Q_2 + P_2 Q_2^2 + P_3 Q_2^3)\frac{(1-\alpha)V_2}{Q_2} \tag{6.18}$$

$$\text{s.t.} \quad H_0 - K_D Q_p^2 = H_w + (K_1 + K_2 \frac{\alpha V_2}{\alpha V_2 + V_1}^2) Q_p^2 \tag{6.19}$$

$$0 \leq \alpha \leq 1 \tag{6.20}$$

Optimal schedule for more than two users: Similarly we can continue to compute optimal solution for three and more users but the number of solution shapes and complexity of conditions increases rapidly. For example, we can assume $2^n - n$ different scenarios that fill the pump capacity and based on the K_is and V_is values one of the scenarios is optimal. So we need to find heuristic or approximate solution. In a heuristic way we recursively use two users solution to determine the optimal scenario.

$$C_k := \sum_{j=k}^{N} V_j \tag{6.21}$$

$$Kpre_i := \sum_{j=1}^{i} K_j \tag{6.22}$$

$$Kpost_i := \sum_{j=i+1}^{N} K_j (C_j/C_i + 1)^2 \tag{6.23}$$

Let $\alpha^*(K_1, K_2, V_1, V_2)$ be the solution to optimize equation (6.18) for the given parameters. Then, the algorithm shown in flowchart of Fig. 6.4 is used to find an approximate solution.

6.4.2 Pump with VSD

Optimal schedule for one user: For a single user, we should find the optimum pump speed and for the optimum speed, optimal flow is maximum possible flow due to head-flow constraint.

$$q_1^*(t) = \begin{cases} q_{max} = \sqrt{\frac{H_0 {n^*}/{n_0}^2 - H_w}{K_D + K_1}} & 0 \le t \le V_1/q_{max} \\ 0 & \text{otherwise} \end{cases} \tag{6.24}$$

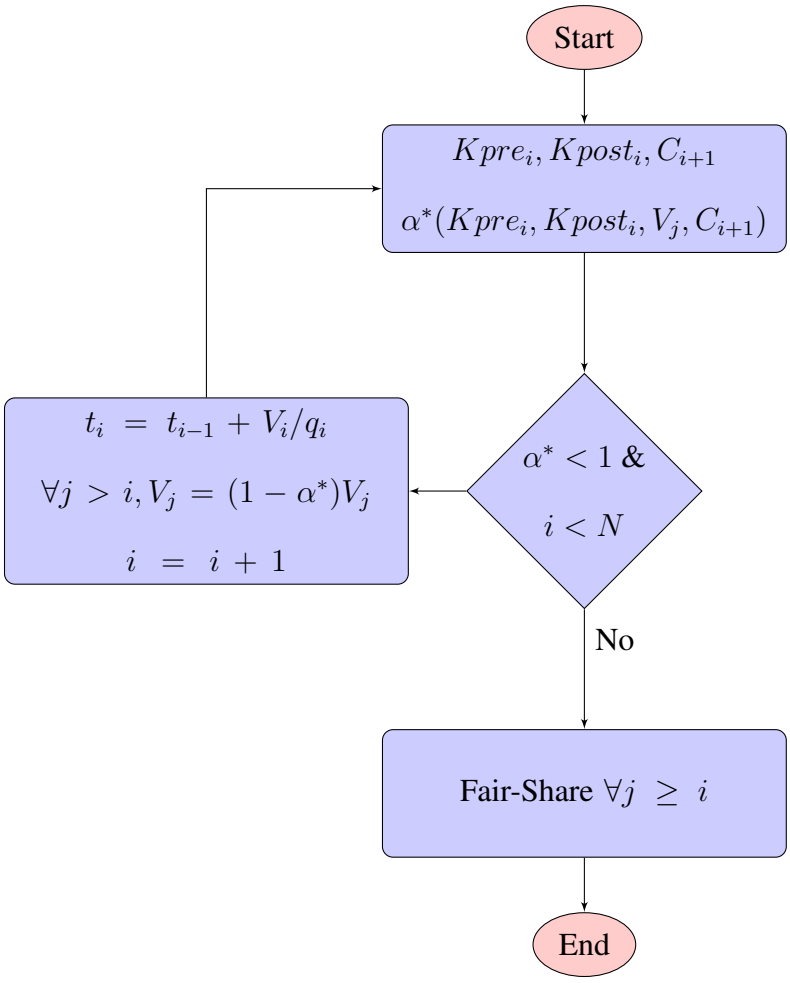

Figure 6.4: Flowchart for more than two users

We find the optimal speed by solving the following nonlinear optimization using Lagrange multipliers.

$$n^* = \underset{n}{\text{argmin}} \frac{V_1}{q_1}(P_0(n/n_0)^3 + P_1(n/n_0)^2 q_1 + P_2(n/n_0)q_1^2 + P_3 q_1^3) \tag{6.25}$$

$$\text{s.t.} \quad H_0(n/n_0)^2 - K_D q_1^2 = H_w + K_1 q_1^2 \tag{6.26}$$

$$q_1 \geq 0 \quad n \leq n_0 \tag{6.27}$$

From optimization of equation (6.25), we find the most efficient operation point which does not depend on demand volume (V_1). V_1 shows itself in time required to finish the task.

Optimal schedule for two users: Similar to fixed speed solution we define a partial share scenario

$$q_1(t) = \begin{cases} q_p \frac{V_1}{V_1 + \alpha V_2} & 0 \leq t \leq \frac{V_1 + \alpha V_2}{q_p} \\ 0 & \text{otherwise} \end{cases} \tag{6.28}$$

$$q_2(t) = \begin{cases} q_p \frac{\alpha V_2}{V_1 + \alpha V_2} & 0 \leq t \leq \frac{V_1 + \alpha V_2}{q_p} \\ q_2^* & \frac{V_1 + \alpha V_2}{q_p} < t \leq \frac{V_1 + \alpha V_2}{q_p} + \frac{V_2(1-\alpha)}{q_2^*} \\ 0 & \text{otherwise} \end{cases} \tag{6.29}$$

where q_p is flow of the pump when both user use the water, α is fraction of V_2 which is shared with first user, q_2^* and n_2^* are solutions of single user case (since it does not depend on volume). Then, the optimization problem becomes:

$$\begin{aligned} \min_{q_p, n_1, \alpha} \quad & BHP(q_p, n_1)\frac{(V_1 + \alpha V_2)}{q_p} + BHP(q_2^*, n_2^*)\frac{(1-\alpha)V_2}{q_2^*} \\ \text{s.t.} \quad & H_0(n_1/n_0)^2 - K_D q_p^2 \geq H_w + (K_1 + K_2 \frac{(\alpha V_2)^2}{(\alpha V_2 + V_1)^2}) q_p^2 \\ & 0 \leq \alpha \leq 1 \end{aligned} \tag{6.30}$$

Even though the optimization of equation (6.30) becomes complex, but we can solve it using KKT conditions. For three users or more we can use same approximation algorithm

of fixed speed pump but the solution of optimization of equation (6.30) is used to determine $\alpha^*(Kpre_i, Kpost_i, V_i, C_{i+1})$.

6.5 Simulation results

In this section we consider 12" pump in Fig.6.1 and we assume static head (water level + minimum required pressure at discharge outlets) is 100 feet. Values of pipe coefficient varies based on the length and diameter of the pipe.

In Fig. 6.5 the required energy for single user versus different pipe coefficient is shown and in Fig.6.6 schedule for using fixed speed and variable speed pump are compared together. The figure clearly shows that VSD pumps perform their tasks more slowly, but with more energy efficiency. VSD pumps also are better for water pipes as they require less pressure and this increases the longevity of the pipes and decreases water leaks. In Fig. 6.7 required energy for pumping 0.5 million gallon of water to each of users based on various pipe coefficient is shown. In both fixed speed and various speed pumps, the partial sharing scenario gives up to 10% improvement with respect to the no share scenario.

6.6 Conclusion

In this chapter, we considered the problem of scheduling water usage for pumping water in a radial water network. For both fixed speed and variable pumps, we derived the exact solution for a two user network. The solution for the two user case is then used to find an approximate solution for networks with more users. Our simulation shows that active participation of users could reduce the energy consumption up to 10%. Finding solution

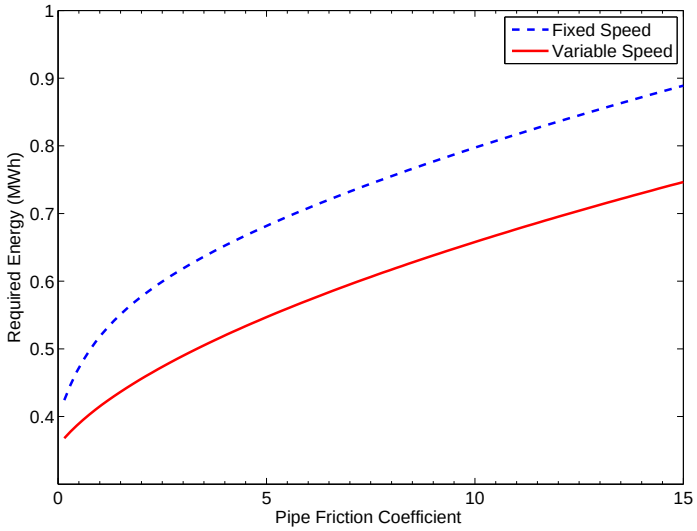

Figure 6.5: Energy required for pumping one million gallons of water using a 12" pump from a 100 feet well

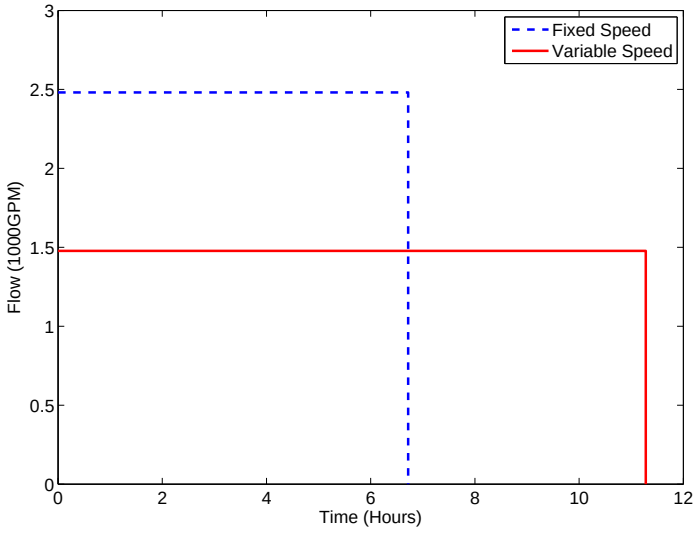

Figure 6.6: Optimal flows for pumping one million gallons to one user ($K_1 = 15$)

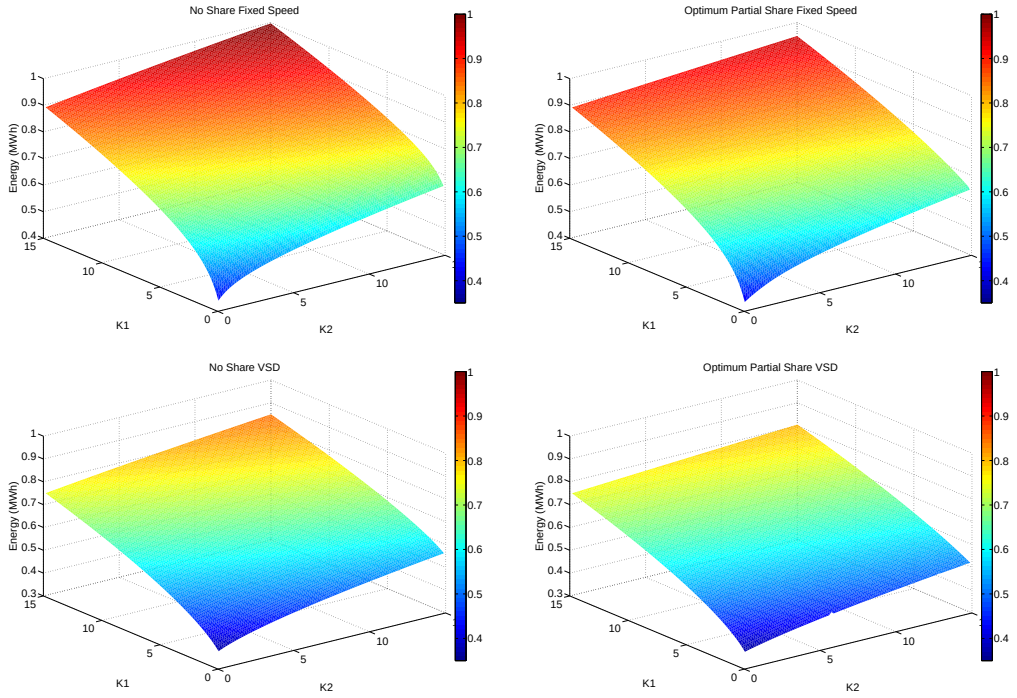

Figure 6.7: Solutions for two users case with various pipe coefficients

for tree networks, performance bounds for the approximate solution of networks with more users, and decentralized algorithm are also interesting but left for future research.

Chapter 7

Summary and Further Dirsctions

Environmental issues such as global warming and climate change have changed people's perception as new more efficient and more sustainable energy sources are being examined and more energy efficient technologies are being developed to reduce energy consumption. To this end this dissertation is mainly related to optimizing energy usage. In Chapter 2-4 we discussed solar forecasting methods which is a requirement for efficient integration of solar generation into the power grid. Significant penetration of intermittent generation like wind and solar in the power grid introduces concerns about reliability of the grid due to stochastic nature of these energy sources. Therefore, in Chapter 5 we discussed reliability of the electric grid using an insurance framework. Finally, an important problem for the energy / water nexus is discussed in Chapter 6. We discussed energy efficient scheduling algorithms for pumping water in water networks.

In the second chapter, we implemented and evaluated simple forecasting methods for predicting solar radiation. Simulation results for three different sites show that normalization using cosine of zenith angle is an effective way for removing both seasonal and daily

effects.

Solar forecasting methods are usually evaluated using symmetric criteria like RMSE or MAE. However, grid operators have more concern about shortage of production rather than its abundance, i.e. overestimation of resources has more serious consequences than underestimation. So in the BA's view the cost function is not symmetric. For this reason we discussed solar radiation forecast under CPWL and LinEx as asymmetric cost functions which are better fitted to the grid operator problem. For each of these cost functions we used two scenarios i.e. adding bias to an unbiased forecast or formulating a directly biased forecast which considers the cost function from the outset. Under CPWL cost the forecast is formulated as a linear program and for LinEx cost we formulated the problem as a convex optimization and solved it by a gradient descent algorithm. Our simulations show that directly biased forecasts have a significant advantage. Simulation results also show that this difference becomes greater as asymmetry in the cost function increases.

We also modified the least mean square (LMS) algorithm according to CPWL and LinEx cost functions to create an online method. A decaying momentum term is used in the learning rule to increase the rate of learning. The proposed online method gives an improvement over batch solutions due to better tracking ability.

We have shown the necessity of using asymmetric cost functions directly in the training phase of simple autoregressive forecast models. More sophisticated learning algorithms based on other methods such as neural networks or wavelets might potentially give better performance, and use of asymmetric cost function when training these models needs further investigation. Here we discussed forecast methods for single sites using their past solar radiation observations. Forecast methods for multiple sites and incorporating exogenous data like weather forecasts are also interesting and left for future research.

Solar forecasting methods in the literature usually use point forecasts which have a single output value for each prediction, but a single value for prediction of a random variable is not enough for many decision makers and more information about uncertainty of future quantity is required. Probabilistic forecasts can give a more complete stochastic characterization using probability distribution over future quantity (for example predicting a CDF function for a random variable). So, two parametric probabilistic forecast methods based on the beta and standard two sided power distributions are presented to predict solar irradiation and their performance are evaluated. A combining procedure using the beta transformed linear opinion pool is utilized to improve the initial forecasts. Our simulations on three different sites show that these methods reduce the uncertainty by providing lower logarithmic score with respect to baseline forecast which uses the marginal distribution. Calibration metrics show that these methods despite their simple structure accurately describe the stochastic characterization of solar irradiation. We obtained this result by using simple zenith angle normalization and tap delay filters to find conditional mean and conditional variance. Many other point forecast methods which are minimizing RMSE or MSE can be used for estimation of conditional mean and variance to obtain a probabilistic forecast based on the approach proposed here.

In addition to this work, some papers have recently discussed probabilistic forecast for solar radiation. However, probabilistic solar forecast is still immature and requires more contributions as combination of different probabilistic forecasts could lead to better a performance. A potential direction is to use support vector conditional density estimation for probabilistic forecast of solar radiation.

Forecasts can be used to make decisions in controlling the electric power grid. In order to make a suitable decision, the influence of intermittent generation on reliability of the

power grid as well as reliability cost and reliability worth should be investigated. To have deeper insights, we discussed challenges related to the reliability of the power grid. System reliability is considered as a public good and a method is required to make the social optimal decision about quantity of reliability level. The commonly used rule of thumb is that outages should happen less often than 1 day in 10 years. However this rule of thumb may be different than the social optimal decision; hence, researchers defined an economic criteria called value of lost load or VOLL which is used to estimate the marginal social cost of unserved loads. VOLL can be used by decision makers to decide on optimal quantity of electricity reliability, however, with current pricing method which bundles the price of electricity and other services, a unique price for this mixed good is not efficient as the VOLL varies for different firms. We proposed an insurance policy as a way to decouple electricity and reliability and a proxy for implementation of benefit taxation which can achieve both optimal quantity and optimal price. The premium of the insurance is optional but depends on two factors, outage probability and losses by the costumers if they experience an outage. The base probability of the outage is computed by the Utility and need to be approved by Public Utilities Commission. Each costumer pay the premium he or she will to pay but the compensation is based on amount of premium divided by the base probability of an outage. Therefore, a risk neutral decision maker will to pay a premium equal to his/her losses times the outage probability. The Utility use part of the collected premium for compensating losses and can use the rest for improving the reliability level more than the baseline. This research is preliminary work where we have introduced ways reliability and insurance can be used in the operation of the electric power grid.

Finally, we considered the problem of scheduling water usage for pumping water in a radial water network. For both fixed speed and variable pumps, we derived the exact

solution for a two user network. The solution for the two user case is then used to find an approximate solution for networks with more users. Our simulation shows that active participation of users could reduce the energy consumption up to 10%. Finding solution for tree networks, performance bounds for the approximate solution of networks with more users, and decentralized algorithm are also interesting but left for future research.

Bibliography

[1] S. E. I. Association *et al.*, "Solar market insight report 2016 year in review," 2017.

[2] M. Rekinger, F. Thies, G. Masson, and S. Orlandi, "Global market outlook for solar power 2015–2019," *SolarPower Europe*, 2015.

[3] X. Luo, J. Wang, M. Dooner, and J. Clarke, "Overview of current development in electrical energy storage technologies and the application potential in power system operation," *Applied Energy*, vol. 137, pp. 511–536, 2015.

[4] R. Bahm, I. S. E. S. A. Section, and I. S. E. S. A. S. S. R. Division, *Satellites and forecasting of solar radiation: proceedings of the First Workshop on Terrestrial Solar Resource Forecasting and on Use of Satellites for Terrestrial Solar Resource Assessment, February 2-5, 1981, Washington*, ser. International Solar Energy Society Series. Publication Office of the American Section of the International Solar Energy Society, 1981, no. pt. 3. [Online]. Available: https://books.google.com/books?id=ieNSAAAAMAAJ

[5] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[6] J. Huang, M. Korolkiewicz, M. Agrawal, and J. Boland, "Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (cards) model," *Solar Energy*, vol. 87, pp. 136–149, 2013.

[7] A. Bracale, P. Caramia, G. Carpinelli, A. R. Di Fazio, and G. Ferruzzi, "A bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control," *Energies*, vol. 6, no. 2, pp. 733–747, 2013.

[8] E. B. Iversen, J. M. Morales, J. K. Møller, and H. Madsen, "Probabilistic forecasts of solar irradiance using stochastic differential equations," *Environmetrics*, vol. 25, no. 3, pp. 152–164, 2014.

[9] M. P. Almeida, O. Perpiñán, and L. Narvarte, "PV power forecast using a nonparametric PV model," *Solar Energy*, vol. 115, pp. 354–368, 2015.

[10] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied energy*, vol. 157, pp. 95–110, 2015.

[11] M. David, F. Ramahatana, P.-J. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models," *Solar Energy*, vol. 133, pp. 55–72, 2016.

[12] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation - with application to solar energy," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3850–3863, 2016.

[13] S. A. Fatemi, A. Kuh, and M. Fripp, "Solar radiation forecast under convex piece-

wise linear cost functions," in *Neural Networks (IJCNN), 2016 International Joint Conference on*. IEEE, 2016, pp. 4985–4990.

[14] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[15] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.

[16] J. Reichl, M. Schmidthaler, and F. Schneider, "The value of supply security: The costs of power outages to Austrian households, firms and the public sector," *Energy Economics*, vol. 36, pp. 256–261, 2013.

[17] P. W. Jowitt and G. Germanopoulos, "Optimal pump scheduling in water-supply networks," *Journal of Water Resources Planning and Management*, vol. 118, no. 4, pp. 406–422, 1992.

[18] H. Zhang, X. Xia, and J. Zhang, "Optimal sizing and operation of pumping systems to achieve energy efficiency and load shifting," *Electric Power Systems Research*, vol. 86, pp. 41–50, 2012.

[19] E. Price and A. Ostfeld, "Iterative linearization scheme for convex nonlinear equations: Application to optimal operation of water distribution systems," *Journal of Water Resources Planning and Management*, 2012.

[20] A. Goryashko and A. Nemirovski, "Robust energy cost optimization of water dis-

tribution system with uncertain demand," *Automation and Remote Control*, vol. 75, no. 10, pp. 1754–1769, 2014.

[21] D. F. Moreira and H. M. Ramos, "Energy cost optimization in a water supply system case study," *Journal of Energy*, vol. 2013, 2013.

[22] V. Puleo, M. Morley, G. Freni, and D. Savić, "Multi-stage linear programming optimization for pump scheduling," *Procedia Engineering*, vol. 70, pp. 1378–1385, 2014.

[23] J. Błaszczyk, A. Karboński, K. Krawczyk, K. Malinowski, and A. Allidina, "Optimal pump scheduling for large scale water transmission system by linear programming," *Journal of Telecommunications and Information Technology*, pp. 91–96, 2012.

[24] J. Nicklow, P. Reed, D. Savic, T. Dessalegne, L. Harrell, A. Chan-Hilton, M. Karamouz, B. Minsker, A. Ostfeld, A. Singh *et al.*, "State of the art for genetic algorithms and beyond in water resources planning and management," *Journal of Water Resources Planning and Management*, vol. 136, no. 4, pp. 412–432, 2009.

[25] M. López-Ibáñez, T. D. Prasad, and B. Paechter, "Ant colony optimization for optimal control of pumps in water distribution networks," *Journal of Water Resources Planning and Management*, 2008.

[26] G. Bonvina, A. Samperiob, C. Le Papeb, V. Mazauricb, S. Demasseya, and N. Maïzia, "A heuristic approach to the water networks pumping scheduling issue," 2015.

[27] R. S. Whaley, R. Hume *et al.*, "An optimization algorithm for looped water networks," in *PSIG Annual Meeting*. Pipeline Simulation Interest Group, 1986.

[28] R. DeMoyer and L. B. Horwitz, *A system approach to water distribution modeling and control*. Lexington Books, 1975.

[29] K. W. Little and B. J. McCrodden, "Minimization of raw water pumping costs using milp," *Journal of Water Resources Planning and Management*, vol. 115, no. 4, pp. 511–522, 1989.

[30] P. Jowitt, R. Garrett, S. Cook, and G. Germanopoulos, "Real-time forecasting and control for water distribution," in *Computer applications in water supply: vol. 2—systems optimization and control*. Research Studies Press Ltd., 1988, pp. 329–355.

[31] L. E. Ormsbee and K. E. Lansey, "Optimal control of water supply pumping systems," *Journal of Water Resources Planning and Management*, 1994.

[32] S. A. Fatemi and A. Kuh, "Solar radiation forecasting using zenith angle," in *IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, 2013, pp. 523–526.

[33] S. Fatemi and A. Kuh, "Solar radiation forecasting under asymmetric cost functions," in *Neural Networks (IJCNN), 2014 International Joint Conference on*, July 2014, pp. 1727–1732.

[34] S. Fatemi, A. Kuh, and M. Fripp, "Online solar radiation forecasting under asymmetrie cost functions," in *Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA)*, Dec 2014, pp. 1–6.

[35] S. A. Fatemi, A. Kuh, and M. Fripp, "Online and batch methods for solar radiation forecast under asymmetric cost functions," *Renewable Energy*, vol. 91, pp. 397–408, 2016.

[36] S. A. Fatemi, A. Kuh, and V. Gupta, "Energy efficient scheduling algorithms for pumping water in radial networks," in *Information Theory and Applications Workshop (ITA), 2016*. IEEE, 2016, pp. 1–6.

[37] S. A. Fatemi, A. Kuh, and M. Fripp, "Probablistic forecast of solar irradiation using combination of two parametric forecasts," *in review*.

[38] M. Z. Jacobson, *Fundamentals of atmospheric modeling*. Cambridge University Press, 2005.

[39] A. Wilcox, S.; Andreas, "Solar resource & meteorological assessment project (solrmap): Rotating shadowband radiometer (rsr)," 2010. [Online]. Available: http://dx.doi.org/10.5439/1052228

[40] R. L. Plackett, "Some theorems in least squares," *Biometrika*, vol. 37, no. 1/2, pp. 149–157, 1950.

[41] M. Ortega-Vazquez and D. Kirschen, "Estimating the spinning reserve requirements in systems with significant wind power generation penetration," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 114–124, Feb. 2009.

[42] M. Fripp, "Greenhouse gas emissions from operating reserves used to backup large-scale wind power," *Environmental science & technology*, vol. 45, no. 21, pp. 9405–9412, 2011.

[43] P. Kundur, *Power System Stability and Control*. New York: Tata McGraw-Hill Education, 1994.

[44] M. Bollen, *The Smart Grid: Adapting the Power System to New Challenges*, ser. Synthesis digital library of engineering and computer science. Morgan & Claypool, 2011.

[45] BAL-STD-002-0, *Operating Reserves*. WECC standard, 2007.

[46] M. de Nooij, C. Koopmans, and C. Bijvoet, "The value of supply security: The costs of power interruptions: Economic input for damage reduction and investment in networks," *Energy Economics*, vol. 29, no. 2, pp. 277 – 295, 2007.

[47] K. H. LaCommare and J. H. Eto, "Cost of power interruptions to electricity consumers in the United States (US)," *Energy*, vol. 31, no. 12, pp. 1845 – 1855, 2006.

[48] Hoyer-Klick *et al.*, "Management and exploitation of solar resource knowledge," in *Proc. 1st Int. Congress on Heating, Cooling and Buildings (EUROSUN 2008),*, 2008.

[49] H. G. Beyer *et al.*, "D 1.1. 3 report on benchmarking of radiation products," *Report under contract*, no. 038665, 2009.

[50] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535 – 576, 2013.

[51] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar

power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017 – 2028, 2012.

[52] E. Lorenz *et al.*, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE J. Sel. Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, March 2009.

[53] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states," *Solar Energy*, vol. 85, no. 5, pp. 967 – 977, 2011.

[54] R. Perez *et al.*, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, vol. 84, no. 12, pp. 2161 – 2172, 2010.

[55] C. W. Chow *et al.*, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881 – 2893, 2011.

[56] J. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," *Renewable Energy*, vol. 52, no. 0, pp. 118 – 127, 2013.

[57] J. Shi *et al.*, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Trans. Industry Applications*, vol. 48, no. 3, pp. 1064–1069, May 2012.

[58] X. Zhang, "A statistical approach for sub-hourly solar radiation reconstruction," *Renewable Energy*, vol. 71, no. 0, pp. 307 – 314, 2014.

[59] S. Al-Alawi and H. Al-Hinai, "An ANN-based approach for predicting global radi-

ation in locations with no direct measurement instrumentation," *Renewable Energy*, vol. 14, no. 14, pp. 199 – 204, 1998.

[60] A. Yona, T. Senjyu, and T. Funabashi, "Application of recurrent neural network to short-term-ahead generating power forecasting for photovoltaic system," in *Power Engineering Society General Meeting, 2007. IEEE*, June 2007, pp. 1–6.

[61] G. Capizzi, C. Napoli, and F. Bonanno, "Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1805–1815, Nov 2012.

[62] S. Jafarzadeh, M. Fadali, and C. Evrenosoglu, "Solar power prediction using interval type-2 TSK modeling," *IEEE Trans. Sustainable Energy*, vol. 4, no. 2, pp. 333–339, April 2013.

[63] S. Chen, H. Gooi, and M. Wang, "Solar radiation forecast based on fuzzy logic and neural networks," *Renewable Energy*, vol. 60, no. 0, pp. 195 – 201, 2013.

[64] A. Yona *et al.*, "Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction," *IEEE Trans. Sustainable Energy*, vol. 4, no. 2, pp. 527–533, April 2013.

[65] R. Aguiar and M. Collares-Pereira, "TAG: A time-dependent, autoregressive, gaussian model for generating synthetic hourly radiation," *Solar Energy*, vol. 49, no. 3, pp. 167 – 174, 1992.

[66] L. Mora-Lpez and M. S. de Cardona, "Multiplicative ARMA models to generate

hourly series of global irradiation," *Solar Energy*, vol. 63, no. 5, pp. 283 – 291, 1998.

[67] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, no. 3, pp. 342 – 349, 2009.

[68] E. Lorenz *et al.*, "Regional PV power prediction for improved grid integration," *Progress in Photovoltaics: Research and Applications*, vol. 19, no. 7, pp. 757–771, 2011.

[69] A. Zellner, "Bayesian estimation and prediction using asymmetric loss functions," *Journal of the American Statistical Association*, vol. 81, no. 394, pp. 446–451, 1986.

[70] C. W. Granger, "Prediction with a generalized cost of error function," *Operations Research Quarterly*, vol. 20, no. 2, pp. 199–207, Jun. 1969.

[71] P. Pinson, C. Chevallier, and G. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *Power Systems, IEEE Transactions on*, vol. 22, no. 3, pp. 1148–1156, Aug 2007.

[72] R. Bessa, V. Miranda, A. Botterud, and J. Wang, "good or bad wind power forecasts: a relative concept," *Wind Energy*, vol. 14, no. 5, pp. 625–636, July 2011.

[73] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.

[74] B. Widrow and S. D. Stearns, "Adaptive signal processing," *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, 491 p.*, vol. 1, 1985.

[75] S. Fischer, A. Kubis, M. Greve, and C. Rehtanz, "Macro-economic calculation of the value of lost load and the costs per hour of blackouts in Germany."

[76] E. Leahy and R. S. Tol, "An estimate of the value of lost load for Ireland," *Energy Policy*, vol. 39, no. 3, pp. 1514–1520, 2011.

[77] T. Zachariadis and A. Poullikkas, "The costs of power outages: A case study from Cyprus," *Energy Policy*, vol. 51, pp. 630–641, 2012.

[78] H. R. Varian, "A bayesian approach to real estate assessment," *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, pp. 195–208, 1975.

[79] K. Eber and D. Corbus, "Hawaii solar integration study," National Renewable Energy Laboratory, Tech. Rep. NREL/TP-5500-57215, 2013.

[80] R. Burden and J. D. Faires, *Numerical analysis*. Nelson Education, 2010.

[81] B. Widrow and E. Walach, "On the statistical efficiency of the LMS algorithm with nonstationary inputs," *IEEE Trans. Information Theory*, vol. 30, no. 2, pp. 211–221, 1984.

[82] L. Bottou, "Online learning and stochastic approximations," *On-line learning in neural networks*, vol. 17, p. 9, 1998.

[83] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural networks*, vol. 12, no. 1, pp. 145–151, 1999.

[84] F. Reis, M. Brito, V. Corregidor, J. Wemans, and G. Sorasio, "Modeling the performance of low concentration photovoltaic systems," *Solar Energy Materials and Solar Cells*, vol. 94, no. 7, pp. 1222 – 1226, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0927024810001376

[85] P. Pinson *et al.*, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[86] M. Holland and K. Ikeda, "Forecasting in wind energy applications with site-adaptive weibull estimation," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 2184–2188.

[87] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, pp. 125–151, 2014.

[88] R. Beckman and G. Tiet jen, "Maximum likelihood estimation for the beta distribution," *Journal of Statistical Computation and Simulation*, vol. 7, no. 3-4, pp. 253–258, 1978.

[89] S. Kotz, J. RenÃ *et al.*, *Beyond beta: other continuous families of distributions with bounded support and applications*. World Scientific, 2004.

[90] M. Stone *et al.*, "The opinion pool," *The Annals of Mathematical Statistics*, vol. 32, no. 4, pp. 1339–1342, 1961.

[91] T. Gneiting and R. Ranjan, "Combining predictive distributions," *Electronic Journal of Statistics*, vol. 7, pp. 1747–1782, 2013.

[92] S. C. Hora, "Probability judgments for continuous quantities: Linear combinations and calibration," *Management Science*, vol. 50, no. 5, pp. 597–604, 2004.

[93] A. Atputharajah and T. K. Saha, "Power system blackouts-literature review," in *2009 International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2009, pp. 460–465.

[94] F. staffs, "Arizona-Southern California outages on September 8 2011 causes and recommendations," Federal Energy Regulatory Commission-North American Electric Reliability Corporation, Tech. Rep., 2012.

[95] P. Pourbeik, P. S. Kundur, and C. W. Taylor, "The anatomy of a power grid blackout," *IEEE Power and Energy Magazine*, vol. 4, no. 5, pp. 22–29, 2006.

[96] R. Billinton and W. Li, *Reliability assessment of electric power systems using Monte Carlo methods*. Springer Science & Business Media, 2013.

[97] M. de Nooij, C. Koopmans, and C. Bijvoet, "The value of supply security: The costs of power interruptions: Economic input for damage reduction and investment in networks," *Energy Economics*, vol. 29, no. 2, pp. 277 – 295, 2007.

[98] A. J. Praktiknjo, A. Hähnel, and G. Erdmann, "Assessing energy supply security: Outage costs in private households," *Energy Policy*, vol. 39, no. 12, pp. 7825–7833, 2011.

[99] K. Kariuki and R. N. Allan, "Evaluation of reliability worth and value of lost load," *IEE proceedings-Generation, transmission and distribution*, vol. 143, no. 2, pp. 171–180, 1996.

[100] K. Kariuki and R. Allan, "Factors affecting customer outage costs due to electric service interruptions," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 143, no. 6, pp. 521–528, 1996.

[101] M. Sullivan, D. Keane, E. P. R. Institute, and S. . C. Freeman, *Outage*

Cost Estimation Guidebook. Electric Power Research Institute, 1995. [Online]. Available: https://books.google.com/books?id=Fo3NHAAACAAJ

[102] D. Kirschen and G. Strbac, "Why investments do not prevent blackouts," *The Electricity Journal*, vol. 17, no. 2, pp. 29–36, 2004.

[103] N. D. Reppen, "Increasing utilization of the transmission grid requires new reliability criteria and comprehensive reliability assessment," in *Probabilistic Methods Applied to Power Systems, 2004 International Conference on*. IEEE, 2004, pp. 933–938.

[104] R. Standler, "liability of electric utility in the USA for outage or blackout," *51pp. http://www. rbs2. com/outage. pdf [accessed July 16, 2014]*, 2011.

[105] T. J. Brennan, "Holding distribution utilities liable for outage costs," *Energy Economics*, vol. 48, pp. 89–96, 2015.

[106] Z. Yang and H. Børsting, "Optimal scheduling and control of a multi-pump boosting system," in *Control Applications (CCA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2071–2076.

[107] A. J. Stepanoff, *Centrifugal and Axial Flow Pumps*. New York: John Wiley & Sons, 1957.

[108] V. S. Lobanoff and R. R. Ross, *Centrifugal Pumps: Design and Application: Design and Application*. Elsevier, 2013.

[109] G. O. Brown, "The history of the Darcy-Weisbach equation for pipe flow resistance," *Environmental and Water Resources History*, vol. 38, no. 7, pp. 34–43, 2002.