Short-Term Prediction Model to Maximize Renewable Energy Usage in Cloud Data Centers

Atefeh Khosravi and **Rajkumar Buyya**

Cloud Computing and Distributed Systems (CLOUDS) Laboratory,
School of Computing and Information Systems,
The University of Melbourne, Australia
atefehk@student.unimelb.edu.au, rbuyya@unimelb.edu.au

Abstract

The increasing demand for services offered by cloud providers results in a large amount of electricity usage by their data center sites and a high impact on the environment. This has motivated many cloud providers to move towards using on-site renewable energy sources to partially power their data centers using sustainable sources. This way, they can reduce their reliance on brown electricity delivered by off-site providers, which is typically drawn from polluting sources. However, most sources of renewable energy are intermittent and their availability changes over time. Therefore, having short-term prediction helps the cloud provider to make informed decisions and migrate the virtual machines (VMs) between data center sites in the absence of the renewable energy. In this chapter, we propose a short-term prediction model using Gaussian mixture model (GMM). The model uses the previously observed energy levels to train itself and predict the energy level for many-steps ahead into the future. We analyzed the accuracy of the proposed prediction model using real meteorological data. The experiment results show that the GMM model can predict up to 15 minutes ahead into the future with nearly 98% accuracy around ±10% of the actual values. This helps the cloud provider to perform online VM migration with performance close to the optimal offline algorithm, which has the full knowledge of renewable energy level in the system. Moreover, the accuracy of the model has been verified using the workload data from Amazon biggest region in US East (N. Virginia). However, due to the confidentiality of that data set, we only rely on the results of the carried experiments using real meteorological renewable energy traces.

Keywords— Cloud computing, Green computing, Renewable energy, Data center, Renewable Prediction

1 Introduction

Cloud computing is a paradigm focused on the realization and long held dream of delivering computing as a utility [1]. It enables businesses and developers access to hardware resources and infrastructure anytime and anywhere they want. Nowadays, the number of individuals and organizations shifting their workload to cloud data centers is growing more than ever. Cloud services are delivered by data center sites each containing tens of thousands of servers, which are distributed across geographical locations. The geographical diversity of computing resources brings several benefits, such as high availability, effective disaster recovery, uniform access to users in different regions, and access to different energy sources.

Over the recent years the use of services offered by cloud computing systems has been increased and different definitions for cloud computing have been proposed. According to the definition by the National Institute of Standards and Technology (NIST) [2]: "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Cloud addresses the issue of under provisioning of resources for a running service and lose the potential users at the peak times or even over provisioning of resources that leads to wastage of capital costs. This definition highlights a major feature for cloud computing that is called elasticity of resources. By delivering computing as a utility to users and providing the resources based on the users' request, the users will be charged on a pay-as-you-go manner, such as other utility pricing models (e.g., electricity and water). In other words, users need not pay any upfront cost and the billing will be based on the usage (e.g. hourly) of the cloud resources.

Cloud delivers three main services to users as shown in Figure 1 and discussed in the following.

- Software as a Service: At the highest level there is Software as a Service (SaaS). SaaS service model, which is an old idea of cloud computing delivers on-demand software to users. Google Apps [3] and Salesforce [4] are examples of services offered in SaaS model. In this model, the control, support, and maintenance of the hardware, platform, and software of the cloud environment is shifted from the end-user to the cloud provider.

- Platform as a Service: Platform as a Service (PaaS) provides computing platform with pre-installed operating system, in order to enable the developers create their own software. By using PaaS, the developers need not concern about the underlying hardware and the operating system. Users can have scalable resources anytime and anywhere. Google App Engine [5], Microsoft Azure[6], and Manjrasoft Aneka [7] are examples of PaaS environment.

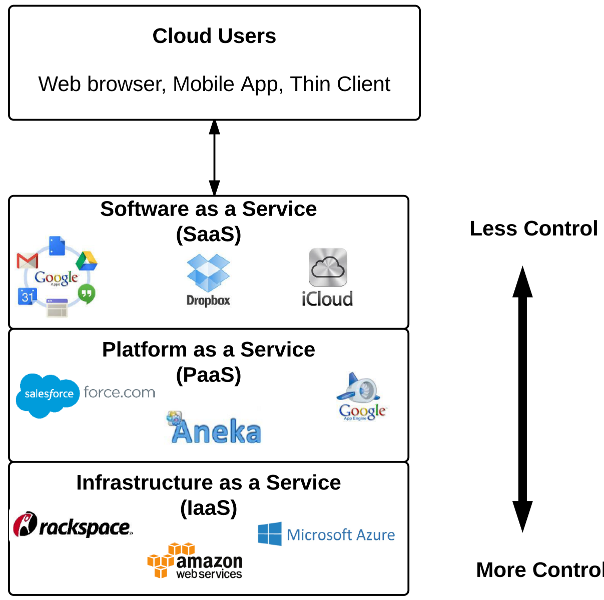

Figure 1: Cloud computing services.

- Infrastructure as a Service: Infrastructure as a Service (IaaS) located at the lowest layer of the cloud service stack offers computing physical resources such as servers, storage, hardware, networking, and virtual machines (VMs) to users. In this model, users have control over the operating system, storage, and applications while they need not manage the underlying infrastructure. Amazon EC2 [8], Google Cloud [9], and Rackspace [10] are some of the well-known IaaS providers.

Services offered by cloud computing are delivered by data centers distributed across the world. One major issue with these data centers is that they are energy intensive, which makes them responsible for 2% of the world's total CO_2 emission [11]. To overcome the problem of high energy consumption and environmental concerns due to the high CO_2 emission of energy sources, there are possible solutions such as improving the data center's efficiency or replacing the polluting (brown) energy sources with clean energy sources. By making data centers aware of energy sources and better utilizing renewable energy, cloud providers are able to reduce the energy consumption and carbon footprint significantly [12].

Further in this section, we elaborate more on our motivation, which is one of the biggest challenges a cloud provider faces, high energy consumption and carbon footprint, and the need to better utilize renewable energy sources.

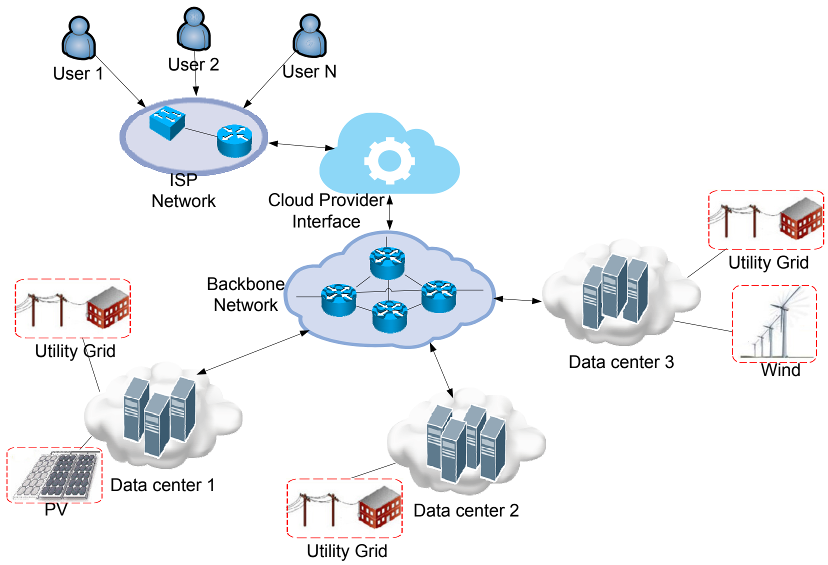

Figure 2: Cloud provider with distributed data center sites with different energy sources.

1.1 Motivation

Data centers are the backbone of the Internet that consist of thousands of servers. They are one of the fastest growing industries that offer different types of services to users around the world. However, data centers are known to consume huge amount of electricity. According to a report by NRDC [13], US data centers in 2013 alone consumed 91 billion kilowatt-hours of electricity. This is equivalent to two-year power consumption of New York City's households and by 2020 is estimated to increase to 140 billion kilowatt hours. This could be equivalent to nearly 150 million tons of carbon pollution. Therefore, many cloud service providers focused on reducing their reliance on electricity driven from fossil fuels and transition to renewable energy sources.

Recently, large cloud providers started building their on-site renewable energy sources. Companies, such as Amazon [14], Facebook [15, 16], Google [17], and Microsoft [18] all have their own on-site solar/wind farms. Renewable energy sources have intermittent nature. This means that their availability changes during the day and based on time of the year. However, since all the large cloud providers have geographically distributed data center sites, as depicted in Figure 2, they can benefit from this location diversity. This helps them to migrate the user requests (e.g., VMs) in the absence of renewable energy in a data center to a site with excess renewable energy.

Since, most sources of renewable energy have intermittent nature knowing the future level of energy helps the cloud provider to make informed decision on when to migrate the VMs to maximize renewable energy usage. The cloud provider can benefit from short-term prediction of renewable energy to perform future-aware online algorithms to migrate the VMs, as it has been stated in our previous work [19]. This helps the provider to increase the performance of the online algorithms close to the optimal offline, which has full knowledge of the future level of renewable energy.

In this chapter, we propose a short-term prediction model based on the Gaussian mixture model [20]. The proposed model predicts renewable energy level for many-steps ahead into the future. A primary requirement to perform prediction is knowing the current and previous states of the renewable energy levels, since the future level can be inferred from current and previous states and their correlation. The GMM model uses history data to train itself. We use renewable energy measurements reported by NREL [21], that have been used in our previous work [19] as real meteorological data, as history and test data in our experiments. Moreover, we verified the accuracy of the proposed prediction model using workload demand collected from AWS biggest region, US East, Virginia. However, due to the confidentiality of that data set, we only rely on the analysis carried out using renewable energy traces collected from NREL.

The rest of the chapter is organized as follows: Section 2 describes the prediction model objective. The formulation and component estimation of the prediction model is explained in Section 3. Section 4 elaborates on the required steps to construct the model. The approaches and methodologies to train the history data is explained in Section 5. Experiment results are presented in Section 6 and Section 7 provides a summary of the chapter.

2 Prediction Model Objective

Energy production at a data center within time period $[1, T]$ is time-series data and can be shown as $\mathbf{y} = [y_1, y_2, .., y_T]^T$, where y_t is the energy production at time t. We show the predicted renewable energy production in a data center at time t as $\hat{y}_t$. The closer the predicted energy $\hat{y}_t$ is to the observed production energy y_t, the more accurate the prediction.

Therefore, our objective is to minimize the prediction error over time interval $[t_1, t_2]$ where $t_1 \leq t_2$, and is stated as follows:

$$\begin{aligned} \underset{\hat{y}_t}{\text{minimize}} \quad & \sum_{t\in[t_1,t_2]} e[(\hat{y}_t - y_t)]\,, \\ \text{subject to} \quad & \hat{y}_t \geq 0\,, \\ \text{and} \quad & predictionModelCost \leq ThresholdCost\,. \end{aligned} \tag{1}$$

The first constraint guaranties the predicted energy production always has non-negative value. Finally, the second constraint guaranties the computation cost of running the prediction, in terms of running time, CPU, and memory usage, over a certain time period will not exceed a predetermined threshold.

3 Prediction Model Formulation

We use the current and previous states of the energy production to perform prediction. The next state of energy production has strong but not deterministic relationship with the current and previous states. This relationship could be shown as a conditional probability. If we denote the current state of the energy production as y_t then the probability of the next state can be denoted as:

$$p(y_{t+1}|y_t, y_{t-1}, ..., y_{t-N+1})\,, \tag{2}$$

where N is considered as the number of previous states taken into account for the prediction. For the sake of simplicity, we show the previous states considered in the prediction as $\mathbf{x} = [y_t, y_{t-1}, ..., y_{t-N+1}]^T$. Therefore, to obtain the energy production prediction we need to compute the following conditional estimation:

$$\hat{y}_{t+1} = \mathbb{E}[y_{t+1}|\mathbf{x}]\,. \tag{3}$$

3.1 Prediction Using Gaussian Mixture Model

To perform the prediction in near future using historical renewable energy production, we use Gaussian mixture models (GMM). In order to obtain the prediction value, first we need to compute $p(y_{t+1}|\mathbf{x})$. Since the aforementioned probability is unknown, we use GMM to approximate it, assuming it is a combination of multiple Gaussian components [20]. GMM is a powerful tool for data analysis and is characterized by M number of mixtures/components, each with a given mean $\boldsymbol{\mu}$, variance $\boldsymbol{\Sigma}$, and weight ω. The GMM probability density function can be written as follows:

$$p(\mathbf{x}|\Theta) = \sum_{j=1}^{M} \omega_j \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\,, \tag{4}$$

where

$$\begin{aligned}
&\Theta = \{(\omega_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), (\omega_2, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), ...(\omega_M, \boldsymbol{\mu}_M, \boldsymbol{\Sigma}_M)\}\,,\\
&\sum_{j=1}^{M} \omega_j = 1\,,\\
&\mathcal{N}(\mathbf{x}, ; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{\boldsymbol{\Sigma}_j \sqrt{2\pi}} e^{-\dfrac{(\mathbf{x} - \boldsymbol{\mu}_j)^2}{2\boldsymbol{\Sigma}_j}}\,.
\end{aligned} \tag{5}$$

GMM parameters, Θ, can be estimated using the expectation-maximization (EM) algorithm [22]. EM is the most popular approach being used and it iteratively optimizes the model using maximum likelihood maximization.

As we mentioned before, the next energy production value has a conditional probability with the current and previously observed production:

$$\begin{aligned}\hat{y} &= \mathbb{E}[y|\mathbf{x}] \\ &= \int yp(y|\mathbf{x})dy\,.\end{aligned} \tag{6}$$

Since $p(y|\mathbf{x})$ in the Equation (6) is not known, we use Bayes' Theorem for its estimation stated as follows:

$$p(y|x) = \frac{p(y, \mathbf{x})}{p(\mathbf{x})}\,, \tag{7}$$

and the joint probability distribution for y and $\mathbf{x}$, $p(y, \mathbf{x})$, could be derived using GMM. Therefore, Equation (7) could be restated as:

$$\begin{aligned}p(y|\mathbf{x}) =& \frac{\sum_{i=1}^{M} \omega_i \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{i\mathbf{x}^T}, \boldsymbol{\Sigma}_{i\mathbf{x}\mathbf{x}})\mathcal{N}(y; \boldsymbol{\mu}_{iy|\mathbf{x}^T} \boldsymbol{\Sigma}_{iy|\mathbf{x}})}{\sum_{j=1}^{M} \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{j\mathbf{x}^T}, \boldsymbol{\Sigma}_{j\mathbf{x}\mathbf{x}})} \\ =& \sum_{i=1}^{M} \beta_i \mathcal{N}(y; \boldsymbol{\mu}_{iy|\mathbf{x}^T}, \boldsymbol{\Sigma}_{iy|\mathbf{x}})\,,\end{aligned} \tag{8}$$

where

$$\begin{aligned}\beta_i &= \frac{\omega_i \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{i\mathbf{x}^T}, \boldsymbol{\Sigma}_{i\mathbf{x}\mathbf{x}})}{\sum_{j=1}^{M} \omega_j \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{j\mathbf{x}^T}, \boldsymbol{\Sigma}_{j\mathbf{x}\mathbf{x}})}\,, \\ \boldsymbol{\mu}_{iy|\mathbf{x}^T} &= \mu_{iy} - \boldsymbol{\Sigma}_{iy\mathbf{x}} \boldsymbol{\Sigma}_{i\mathbf{x}\mathbf{x}}^{-1} (\boldsymbol{\mu}_{i\mathbf{x}^T} - \mathbf{x})\,.\end{aligned} \tag{9}$$

Finally, by substituting Equation (8) into Equation (6), we have:

$$\begin{aligned}\hat{y} &= \sum_{i=1}^{M} \beta_i \int y\, \mathcal{N}(y; \boldsymbol{\mu}_{iy|\mathbf{x}^T}, \boldsymbol{\Sigma}_{iy|\mathbf{x}^T})dy \\ &= \sum_{i=1}^{M} \beta_i \boldsymbol{\mu}_{iy|\mathbf{x}^T}\,.\end{aligned} \tag{10}$$

3.2 Optimal GMM Components Estimation

We use expectation maximization (EM) algorithm to estimate GMM parameters Θ. EM is an iterative method to find the maximum likelihood estimate (MLE) of the parameters. In order for EM to perform the two steps of expectation (E) and maximization (M), it needs to receive the number of GMM mixtures as an input.

There have been several studies and different methods to obtain the optimal number of mixtures and selecting the efficient model, rather than simply taking a random or educated guess. Bayesian information criterion (BIC) [23] is a criterion introduced for model selection and is penalized based on the model complexity. BIC maximizes the maximum likelihood function for each model. It is based on the increasing function of an error and the model with the lowest BIC, the more efficient in terms of predicting the demand.

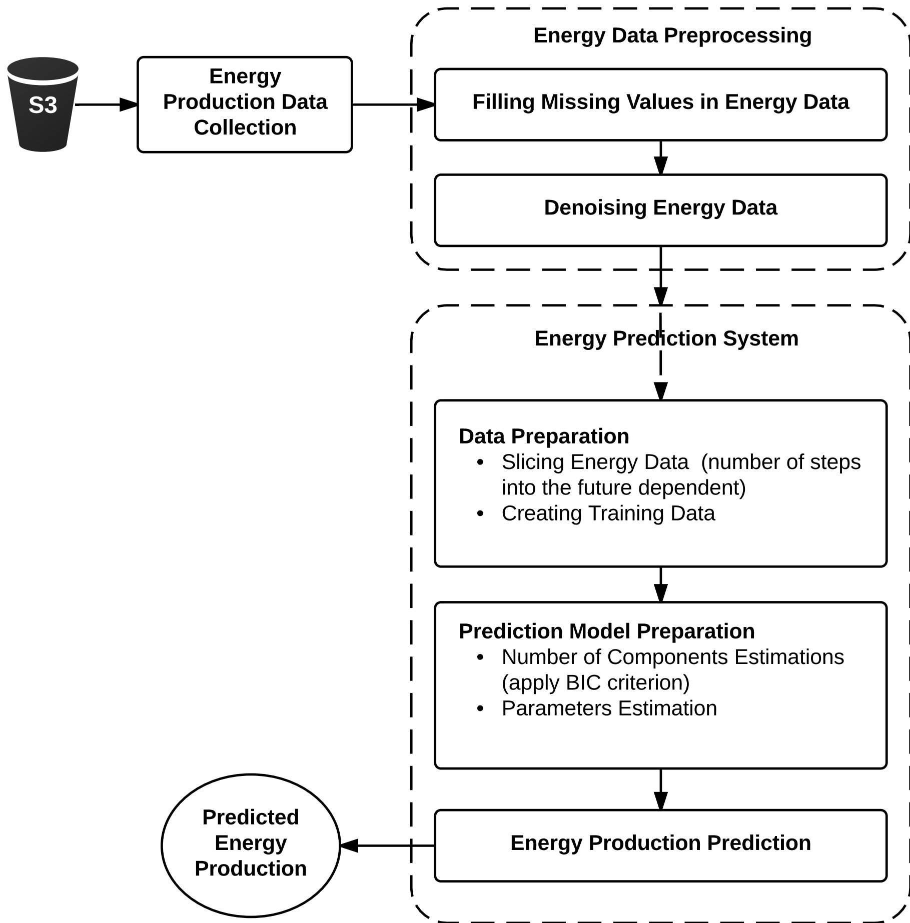

Figure 3: Renewable energy production prediction model.

4 Construction of Prediction Model

Figure 3 shows the required steps towards constructing the prediction model. Different steps involved in performing the prediction are discussed in the rest of this section.

4.1 Filling Missing Values in Renewable Energy History Data

Access to accurate history data is critical for prediction. Since having access to perfect history data is not always the case, often there are missing points in time regarding collected history data. Keeping the time-stamp related to each renewable energy data is important to feed into the prediction model. Filling the gaps by simply shifting the energy history data back in time changes the energy data-time mapping. Therefore, we need to fill-up the missing values in the collected energy data while keeping each renewable energy's time-stamp. For each collected solar and wind energy, if there are missing data points in the beginning or at the end of a time period, we replicate the first or last observed energy data, respectively. Otherwise, if there are missing energy data in the middle of the time series, we use linear interpolation between the first and the last observed energy data. As presented in Figure 3, filling missing values in the renewable energy history data is part of the preprocessing step, before performing the prediction.

4.2 Denoising the Renewable Energy Data

Before training the data and performing the prediction, we need to smooth the collected renewable energy data and remove the sharp acceleration and deceleration of the energy data to achieve a fair prediction. To smooth the history data, we use the fast fourier transform (FFT) algorithm [24] to remove the high frequencies in the energy data and reconstruct it again with only low frequency information.

4.3 Training History Data

As shown in Figure 3, we need to prepare the history data to feed into the prediction model. Training set will be constructed according to the following pattern.

$$\mathbf{Z} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_N & y_1 \\ x_2 & x_3 & x_4 & \dots & x_{N+1} & y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_T & x_{T+1} & x_{T+2} & \dots & x_{T+N} & y_T \end{bmatrix}, \tag{11}$$

where, in our model, $y_i = x_{N+i}$ for $i \in [1, T]$.

To perform the renewable energy production prediction $\hat{y}_{T+1}$, we use the previously observed production values. The granularity of the data history should be equal to the length of the prediction being performed from the last observed renewable energy upto 1-step ahead in time. We denote the granularity of the data history as g, which should be equal to performing the prediction for 1-step ahead into the future (g = 1-step ahead prediction length).

4.4 Feature Set Selection

Performing renewable energy prediction requires access to the history data and training the data to estimate prediction model parameters. As stated earlier, we use N previously observed states to predict the next energy production. GMM parameters estimation are driven from running $gmm.fit$ on the training set $\mathbf{Z}$ containing history data. The training set is constructed from multiple rows, each equal to a $\mathbf{z} = [\mathbf{x}, x_{t+1}]$ vector, where $\mathbf{x} = [x_{t-N+1}, ..x_{t-1}, x_t]$. The elements of $\mathbf{z}$ do not necessarily need to be consecutive observed values. Vector $\mathbf{z}$ elements selection have a major effect on the estimation of the prediction model parameters and accordingly the predicted value of the energy production.

5 Prediction Approach and Methodologies

As we mentioned earlier, training the data and filling the training matrix with the right feature set is important to lead us to an accurate prediction. Depending on the time-step ahead into the future that the prediction is taking place, we consider two different approaches to train the data history. To perform the energy prediction for the s^{th}-step ahead into the future, the following two approaches are considered:

- Short-term approach: Selecting every subsequent s^{th} item in the data history.

- Long-term approach: Selecting every subsequent hour:minute corresponding to the hour:minute of the s^{th}-step in the data history.

Moreover, in order to construct the training matrix we consider two different methodologies, as

- Direct multi-step ahead prediction: Direct multi-step ahead prediction (DMSA) performs the energy prediction for s-steps ahead into the future using only the history data. In this approach, the energy production prediction for $\hat{y}_{t+s}$ is independent of the prediction results for energy production before time $t + s$ and is made directly using the data available upto time t.

- Propagated multi-step ahead prediction: Propagated multi-step ahead (PMSA) prediction uses the predicted energy production as an input to the model for next energy production prediction. PMSA uses the $\hat{y}_{t+s-1}$ value as an input to predict the value of $\hat{y}_{t+s}$. The main aim of propagated prediction is to use the results of the previous successful predictions for the next predictions, since prediction results are more accurate for time-steps closer to the last observed energy data.

6 Prediction Model Evaluation

This section discusses the experiment setup and the validation of the prediction model. However, as it has been stated before, the accuracy of the prediction model has been tested using the workload demand collected from AWS biggest region, US East, Virginia. We used one month of data with granularity 15 minutes as history data to train the model and predict 7 days ahead into the future. However, due to the confidentiality of the used data-set and also our goal to validate the model for renewable energy production, we run a separate set of experiments based on renewable energy production prediction.

6.1 Experiment Setup

6.1.1 Renewable Energy Traces

We use the renewable energy measurements from NREL [21] to calculate solar and wind energy production for a data center. The solar and wind energy traces used in this chapter are the same as the renewable energy used in our previous work [19]. The measurements are with 1 minute granularity from May, 1st to May, 29th 2013. We use Global horizontal irradiance (GHI) measurements to calculate the output of the solar photovoltaic (PV). The GHI measurements are for PV flat panels on tilted surface at a 45-degree angle and PV efficiency of 30%. We calculate the solar output based on [25] and the total area for the flat plates is considered to be $100m^2$, derived from the configuration by Solarbayer [26].

To calculate wind energy production, we use the proposed model by Fripp et al. [27]. We feed the wind speed, air temperature, and air pressure, derived from NREL measurements, to the model to calculate wind power at the data center, assuming the data center uses a GE 1.5MW wind turbine.

6.1.2 Benchmark Prediction Models

We compare the results of the prediction model against three different models. Naive that assumes prediction at each point in time is the same as the previously observed value, $\hat{y_{t+1}} = y_t$, linear regression [28] and random forest [29].

6.2 Prediction Analysis Metrics

We investigate the performance of the prediction model by studying the following quality metrics:

6.2.1 Bounded Predicted Values

We use bounded predicted values as a measure to quantify the percentage of the predicted values around $x\%$ of the actual values. This is a good measurement to know for different prediction models, what is the percentage of the predicted values bounded within an error margin (e.g., $\pm 20\%$).

6.2.2 R-Squared

In analyzing the accuracy of a prediction, a good prediction model would have the predicted versus actual values as close to the 45-degree line, as shown in the Figure 5. R-squared is a statistical measure that shows how close the predicted values are to the actual values. R^2 gives an intuitive measure of the proportion of the predicted values that could be explained by the actual values. In other words, an R^2 with value x means that $x\%$ of the prediction variation is explained by the actual values.

R^2 value is between 0 and 100%. The higher the R-squared, the better the prediction fits the actual values. If a prediction model could explain 100% of the variance, the predicted values would always equal the actual values and therefore, all the data points would fall on the 45-degree line.

6.2.3 Standard Error

Standard error (S), same as R^2, tells us how well the predicted and actual values would fall on the same line. Standard error is the average distance between the predicted and the actual values. The smaller the S the better the prediction and indicates that the predicted and actual values fall on the 45-degree line. Moreover, standard error is a good indication to show the accuracy of the prediction. A standard error with value s tells that approximately 95% of the predicted versus actual values fall within $\pm 2 \times s$ of the 45-degree line.

6.2.4 Mean Absolute Error (MAE)

Using a metric that measures the average magnitude of the errors is always useful and indicates how big of an error can be expected from the prediction on average. A perfect prediction would have a MAE zero. Since MAE is skewed in favor of large errors (prediction outliers), we need to use other metrics, such as p^{th}-percentile to better validate the accuracy of the prediction model.

6.2.5 P-Percentile

P-percentiles are useful to know the distribution of the prediction error. A p^{th} percentile of a distribution shows that roughly $p\%$ of the error values are equal to or less and $(1-p)\%$ of the error values are larger than that number. Percentiles range in $[0, 100]$. The 0^{th}-percentile shows the min and 100^{th}-percentile shows the max value in a distribution. We measure the p^{th} percentiles on the absolute values of the prediction error ($|\hat{y}-y|$). This way we focus on the unsigned errors and measure how close the prediction and actual values are together, without considering the direction of the error.

It should be noted that when reporting percentiles, we need to consider that if the data distribution is heavy-tailed (right-skewed), significant outliers could be hidden, even not reflected in 90^{th} or 99^{th} percentiles. Therefore, we also report p-100 which shows the maximum error value in the prediction.

6.3 Prediction Results and Analysis

In the following, we validate the accuracy of the proposed prediction model using the renewable energy measurements from NREL [21]. From the collected renewable energy levels for May 2013, we consider the first three weeks as the data history to train the model and the last 8 days as test data to verify the prediction accuracy. We run the prediction model on the previously observed

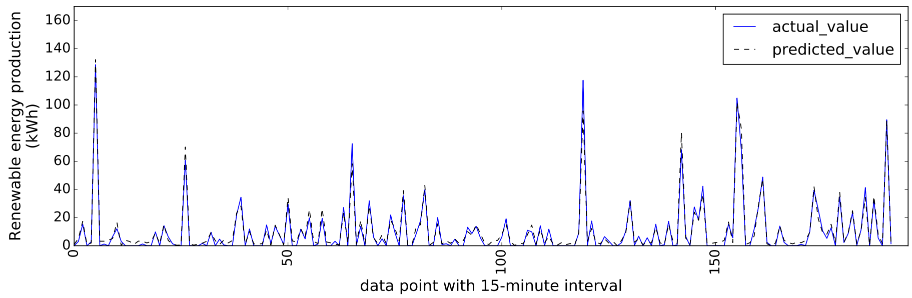

Figure 4: Results of prediction model for 8 days period of renewable energy production for 15-minute ahead prediction.

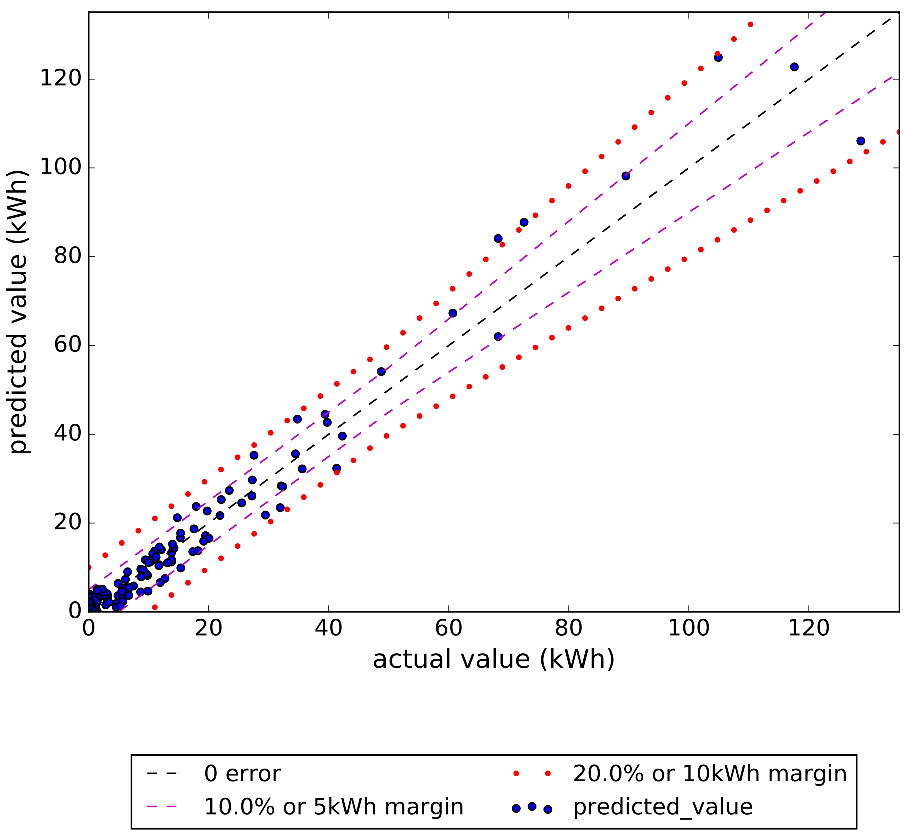

Figure 5: Predicted vs. actual values for 8 days period of renewable energy production for 15-minute ahead prediction with ±10% and ±20% around the actual value.

renewable energy production (the data history) to predict the renewable energy level for the next 15 minutes with the granularity of 1 minute. Then, we move the data history window 15-minute ahead to predict the next 15 minutes. We repeat this till we predict 8 days of renewable energy level.

Since the prediction window size is relatively small, 15 minutes, we use the short-term approach, discussed in Section 5, to fill the elements in the training matrix. Moreover, we use DMSA methodology, which is independent from the newly predicted values, for feature set selection and training the data. We defer applying long-term approach and PMSA methodology for the interested reader. However, in the carried experiments using AWS data to perform one-week ahead prediction, we applied short-term approach for predictions up to 36 hours ahead in time and long-term approach beyond that time.

Figure 4 shows the renewable energy production prediction against the actual values for 8 days. We also demonstrate the GMM prediction results using scatter plot, Figure 5. As it can be seen in this figure, we measure the percentage of the predicted values bounded within ±10% and ±20% relative error and each with considering an absolute error of 5kWh and 10kWh, respectively. Using an absolute error constraint on top of the relative error margins, prevents small errors to affect our decision making. This has been stated as bounded predicted values in Table 1.

We use the previously discussed prediction analysis metrics to evaluate the accuracy of the prediction. The results are presented in Table 1. The prediction model column states GMM model and other benchmark models used in our analysis. The results show that almost all the predicted values in GMM (≈ 100%), fall within 20% of observed actual values, whilst linear regression, random forest and naive model all are lower than GMM. Even checking bounded predictions bounded within ±10% of the actual values is still close to perfect prediction (97.39%). This means GMM can predict renewable energy with considerably high precision almost similar to real time measuring.

In the rest of the reported metrics, R^2, S, MAE, P-90, P-99, and P-100, GMM is performing better than the rest of the models. Having R^2 of 97% shows that almost all the predicted values are aligned and could be explained by the actual values. Moreover, as per the measured MAE, the prediction error on average is 2.42 kWh, which is a negligible value.

Table 1: Prediction accuracy under different quality metrics.

Prediction Model	*Bounded Predictions*	R^2	S	MAE	P-90	P-99	P-100
GMM	99.48%	97%	0.18	2.42	4.16	4.39	21.76
Linear regression	89.81%	86.34%	0.21	3.97	6.78	8.01	25.72
Random forest	81.27%	77.71%	0.43	5.45	9.63	11.84	31.65
Naive	47.41%	0.01%	1.7	16.04	35.56	118.34	128.67

7 Summary

In this chapter, we presented a short-term renewable energy production prediction to predict the renewable energy level for many time-steps ahead into the future. The proposed model is based on the Gaussian mixture model and uses history data to train itself and predict the next level of renewable energy in a data center. Knowing the future level of renewable energy helps the cloud provider to make an informed decision to migrate the VMs in the absence of the renewable energy in a data center to a data center with excess renewable energy. This way, the cloud provider can maximize the usage of renewable energy.

To validate the accuracy of the proposed model, we used renewable energy measurements by NREL. The prediction results show that GMM model can predict up to 15 minutes ahead into the future with nearly 98% precision around $\pm10\%$ of the actual values. This means that cloud provider can perform online VM migrations with performance close to the optimal offline, that has the full knowledge of the future level of renewable energy.

References

[1] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, and Ivona Brandic. Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation computer systems*, 25(6):599–616, 2009.

[2] Peter Mell and Tim Grance. The nist definition of cloud computing. *National Institute of Standards and Technology*, 53(6):50, 2009.

[3] Google Apps. `https://apps.google.com/`. (accessed on 18/04/2017).

[4] Salesforce. `https://www.salesforce.com/`. (accessed on 18/04/2017).

[5] Google App Engine. `https://appengine.google.com/`. (accessed on 18/04/2017).

[6] Microsoft Azure. `https://azure.microsoft.com/`. (accessed on 18/04/2017).

[7] Christian Vecchiola, Xingchen Chu, and Rajkumar Buyya. Aneka: a software platform for .NET-based cloud computing. *High Speed and Large Scale Scientific Computing*, 18:267–295, 2009.

[8] Jinesh Varia. Best practices in architecting cloud applications in the aws cloud. *Cloud Computing: Principles and Paradigms*, pages 457–490, 2011.

[9] Google Cloud. `https://cloud.google.com/`. (accessed on 18/04/2017).

[10] Rackspace. `https://www.rackspace.com/`. (accessed on 18/04/2017).

[11] Jennifer Mankoff, Robin Kravets, and Eli Blevis. Some computer science issues in creating a sustainable world. *IEEE Computer*, 41(8):102–105, 2008.

[12] James W Smith and Ian Sommerville. Green cloud: A literature review of energy-aware computing. *Dependable Systems Engineering GroupSchool of Computer Science, University of St Andrews, UK*, 2010.

[13] NRDC and Anthesis. Scaling Up Energy Efficiency Across the Data Center Industry : Evaluating Key Drivers and Barriers. 2014.

[14] AWS and Sustainable Energy. `http://aws.amazon.com/about-aws/sustainable-energy/`. (accessed on 13/03/2017).

[15] Facebook Installs Solar Panels at New Data Center. `http://www.datacenterknowledge.com/archives/2011/04/16/facebook-installs-solar-panels-at-new-data-center/`. (accessed on 13/03/2017).

[16] Facebook in Fort Worth: Our newest data center. `https://code.facebook.com/posts/1014459531921764/facebook-in-fort-worth-our-newest-data-center/`. (accessed on 13/03/2017).

[17] Renewable energy. `http://www.google.com/green/energy/`. (accessed on 13/03/2017).

[18] Microsoft To Use Solar Panels in New Data Center. `http://www.datacenterknowledge.com/archives/2008/09/24/microsoft-uses-solar-panels-in-new-data-center/`. (accessed on 13/03/2017).

[19] Atefeh Khosravi, Adel Nadjaran Toosi, and Rajkumar Buyya. Online virtual machine migration for renewable energy usage maximization in geographically distributed cloud data centers. *Concurrency and Computation: Practice and Experience*, 2017. doi:`10.1002/cpe.4125`.

[20] DM Titterington, AFM Smith, and UE Makov. Statistical analysis of finite mixture distributions. 1985.

[21] Measurement and Instrumentation Data Center (MIDC). `http://www.nrel.gov/midc/`. (accessed on 25/01/2017).

[22] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[23] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[24] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[25] Photovoltaic Education Network. `http://pveducation.org/`. (accessed on 25/01/2017).

[26] Solarbayer. `http://www.solarbayer.com/`. (accessed on 25/01/2017).

[27] Matthias Fripp and Ryan H Wiser. Effects of temporal wind patterns on the value of wind-generated electricity in california and the northwest. *IEEE Transactions on Power Systems*, 23(2):477–485, 2008.

[28] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2015.

[29] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.