

家庭電気消費量の典型的パターンの抽出

沈 浩 洋^{†1} 日 野 英 逸^{†1}
村 田 昇^{†1} 若 尾 真 治^{†1}

太陽光発電、風力発電、及びコジェネレーションシステム（熱電併給システム）は、家庭の発電手段として実用化され、導入が進んでいるが、これらの発電手段にはそれぞれ長所と短所がある。実際にこれらの発電手段を導入するには、家庭の電力消費パターンに応じて発電システムを最適に組み合わせる必要がある。本研究では、家庭の電力消費量の典型的なパターンを抽出する手法を提案する。一日の電気消費量を混合正規分布でフィッティングし、一般化 KL 情報量を用いて家庭の電力消費データを一日単位でクラスタリングする。Annex42 データセットを使い、提案手法により電気消費量の典型的なパターンを抽出できることを示す。

Extraction of Basic Patterns of Household Energy Consumption

Solar power, wind power, and co-generation (combined heat and power) systems are possible candidate for household power generation. These systems have their advantages and disadvantages. To propose the optimal combination of the power generation systems, the extraction of basic patterns of energy consumption of the house is required. In this study, energy consumption patterns are modeled by mixtures of Gaussian distributions. Then, using the symmetrized Kullback-Leibler divergence as a distance measure of the distributions, the basic pattern of energy consumption is extracted by means of hierarchical clustering. By an experiment using the Annex 42 dataset, it is shown that the proposed method is able to extract typical energy consumption patterns.

^{†1} 早稲田大学 先進理工学研究所 電気情報生命専攻

School of Science and Engineering, Department of Electrical Engineering and Bioscience, Waseda University

1. はじめに

再生可能エネルギーは太陽光、風力、水力、バイオや地熱などの自然の力を用いたエネルギーの総称である。化石燃料と比べると、これらのエネルギーは作る際に有害物質を放出せず、環境への負荷が少ない発電手法として注目されている。

家庭用の太陽光発電パネル³⁾、小型風力発電機⁴⁾、コジェネレーションシステム⁵⁾などが実用化され、家庭でも発電手段を選べるようになってきている^{1),2)}。身近な家庭でも非枯渇性エネルギーにシフトし、必要に応じて効率よく化石燃料を利用すれば、エネルギー問題に貢献できると考えられる。効率的なエネルギーの利用のためには、エネルギー供給システムの特性の理解と同時に、電力消費パターン⁶⁾の把握が不可欠である。

上記の発電手段にはそれぞれの特徴がある。例えば、太陽光発電パネルは太陽光のみで発電が可能であり、メンテナンスの手間が少ないという利点がある一方、夜に発電できないという欠点がある。小型風力発電機は風さえあればいつでも発電できるという利点があるが、定期的なメンテナンスが必要である。また、太陽光発電も風力発電も自然からエネルギーを作るため、発電量は天気などの要素に左右されやすく、出力が安定ではない。コジェネレーションシステムは電気と熱を同時に安定供給できるが、化石燃料を原料として使う。従って、家庭向け発電システムを導入する前に、その家庭に対し、最適な発電手段の組み合わせを提案する必要があり⁷⁾⁻⁹⁾、その家庭の電気消費量の典型的なパターンを抽出しなければならない。家庭の電気消費量には、家族構成、季節、曜日などによって複数のパターンが存在すると考えられる。本研究は、家庭の電気消費量を解析し、その家庭の典型的な電気消費パターンを抽出することを目的とする。家庭の電気消費量データを時系列データとして扱い、一日のデータに混合正規分布を当てはめる。

2. 電気消費量の典型的パターン

本節では、1日のデータに混合正規分布を当てはめ、EM アルゴリズムを用いてパラメータを推定する手法を説明する。

2.1 混合正規分布

混合分布とは、複数の確率分布を混合したものであり、対象となる確率分布を複数の分布の加重平均によって近似する。混合モデルを使うことで、複雑な消費パターンを少数のパラメータで表すことができる。

代表的な混合分布の一つである混合正規分布では、確率密度関数を正規分布

$$\phi(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right), \quad (1)$$

の加重結合で表す。ここで、 $x \in \mathbb{R}^d$ であり、 $\mu \in \mathbb{R}^d$ 及び $\Sigma \in \mathbb{R}^{d \times d}$ はそれぞれ平均、共分散行列である。行列 A に対して $|A|$ でその行列式を表す。混合正規分布は M 個の重み付きの正規分布の和で表される：

$$p(x; \theta) = \sum_{m=1}^M \pi_m \phi(x; \mu_m, \Sigma_m). \quad (2)$$

また、 π_m は混合比であり、次式を満たす：

$$\sum_{m=1}^M \pi_m = 1, \quad \pi_m \geq 0; (m = 1, \dots, M). \quad (3)$$

各要素となる分布を、添字 m , $m = 1, \dots, M$ で表す。

混合正規分布のパラメータは一般に EM アルゴリズム¹⁰⁾ で推定する。EM アルゴリズムは潜在パラメータを有する確率モデルにおいて、尤度関数が最大になるようにパラメータを求める手法であり、E ステップと M ステップの 2 つのステップで構成されるアルゴリズムである。E ステップでは、現在推定されているパラメータに基づいて、モデルの対数尤度の期待値を計算する。M ステップでは、E ステップで求めた対数尤度の期待値を最大となるようにパラメータを求める。 $\{x_1, \dots, x_n\}$ を入力データとし、以下のようにパラメータを定義する：

$$\theta = \{\pi_m, \mu_m, \Sigma_m\}_{m=1}^M, \quad \pi_m \in \mathbb{R}, \mu_m \in \mathbb{R}^d, \Sigma_m \in \mathbb{R}^{d \times d}.$$

t 回目の繰り返しにおけるパラメータ値を $\theta^{(t)}$ とする。パラメータ θ が収束するまで、以下の EM ステップを繰り返し、パラメータを更新する：

E-step: 尤度関数の近似である Q 関数を計算する：

$$Q(\theta, \theta^{(t)}) = \sum_{i=1}^n \sum_{m=1}^M p(m|x_i; \theta^{(t)}) \log p(x_i, m; \theta). \quad (4)$$

M-step: Q 関数を最大化するための $\theta^{(t+1)}$ を求める：

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta, \theta^{(t)}). \quad (5)$$

より具体的には、混合正規分布の場合、パラメータは次のように更新される¹¹⁾,

$$\pi_m^{(t+1)} = \frac{\sum_{i=1}^n w_i^m}{n}, \quad (6)$$

$$\mu_m^{(t+1)} = \frac{\sum_{i=1}^n w_i^m x_i}{\sum_{i=1}^n w_i^m}, \quad (7)$$

$$\Sigma_m^{(t+1)} = \frac{\sum_{i=1}^n w_i^m x_i x_i^T}{\sum_{i=1}^n w_i^m} - \mu_m^{(t+1)} (\mu_m^{(t+1)})^T. \quad (8)$$

ここで

$$w_i^m = p(m|x_i; \theta^{(t)}) \quad (9)$$

である。

3. クラスタリングによる典型パターンの抽出

本節では、クラスタリングを行うために必要な距離の尺度を導入する。

3.1 KL 情報量

Kullback-Leibler 情報量 (KL 情報量)¹²⁾ は 2 つの確率分布の差を測る尺度である。ただし、KL 情報量は対称性を持たないため、厳密な意味での距離ではないことに注意する。連続分布 p と q があり、 p から q への KL 情報量は次のように定義される：

$$D(p, q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx. \quad (10)$$

KL 情報量の計算は確率分布の推定を伴うので、直接データから計算することが難しい。そこで、KL 情報量を近似手法が提案されている^{13),14)}。まず、二つの確率分布が共に正規分布の場合、KL 情報量はその平均と分散で表される：

$$D_g(p, q) = \frac{1}{2} \left[\log \frac{|\Sigma_q|}{|\Sigma_p|} + \text{Tr}(\Sigma_q^{-1} \Sigma_p) - d + (\mu_p - \mu_q)^T \Sigma_q^{-1} (\mu_p - \mu_q) \right]. \quad (11)$$

ここで、 μ_p と μ_q は正規分布 $p(x)$ と $q(x)$ の平均であり、 Σ_p と Σ_q はそれらの共分散行列である。2 つの混合正規分布 $p(x; \theta)$, $q(x; \theta')$ に対し、 θ と θ' をそれぞれの分布のパラメータとすると、混合正規同士の KL 情報量は次のような近似公式で計算される：

$$D_{gm}(p, q) = \sum_{m=1}^M \pi_m \log \frac{\sum_{m'=1}^M \pi_{m'} \exp^{-D_g(\phi_m, \psi_{m'})}}{\sum_{m'=1}^M \omega_{m'} \exp^{-D_g(\phi_m, \psi_{m'})}}. \quad (12)$$

$\{\pi_m\}_{m=1}^M$ と $\{\omega_{m'}\}_{m'=1}^M$ は混合正規分布の混合比であり、混合正規分布 $p(x; \theta)$ と $q(x; \theta')$ の各成分を $\phi_m(x)$ と $\psi_{m'}(x)$ で定義する。この近似式では $D_g(\phi_m, \psi_{m'})$ は (11) より簡単に計算できる。本研究では、家庭の一日の電力消費パターンに混合正規分布を当てはめるた

めに、一日の総消費電力で全体を割ることで正規化を行った。そのため、通常の KL 情報量では総消費量の情報は失われてしまう。そこで、一日の総消費量を距離に反映させるために、確率分布 p に対応する一日の総消費量 $V(p)$ と q に対応する総消費量 $V(q)$ を導入する。次に正規化を行わない 2 つの混合正規分布を

$$\tilde{p}(x) = V_p \cdot p(x),$$

$$\tilde{q}(x) = V_q \cdot q(x),$$

によって定義し、これらの違いを測る尺度として一般化 KL 情報量¹³⁾ を用いる:

$$D(\tilde{p}, \tilde{q}) = \int_x \left(\tilde{p}(x) \log \frac{\tilde{p}(x)}{\tilde{q}(x)} - \tilde{p}(x) + \tilde{q}(x) \right) dx, \\ \simeq V_p D_{gm}(p, q) + V_p \log \frac{V_p}{V_q} - V_p + V_q.$$

この近似式では $D_{gm}(p, q)$ は (12) より簡単に計算できる。

3.2 階層的クラスタリング

次に、クラスタリングの手法によって、フィッティングしたパラメータで特徴付けられる一日の電力消費量データを、少数の典型的パターンに分類する^{15),16)}。クラスター分析には、主に階層的クラスタリングと非階層的クラスタリングの 2 種類がある。階層的クラスタリングは柔軟にクラスターの数を変えることができるため、本研究では、前者の階層的クラスタリング手法を用いる。

クラスタリングにおいて最も重要な要素の一つは、2 つの要素の近さを測る距離の決め方である。ここでは、前節で導入した一般化 KL 情報量¹⁷⁾ を対称化して用いる:

$$d_{KL}(\tilde{p}, \tilde{q}) = \frac{D(\tilde{p}, \tilde{q}) + D(\tilde{q}, \tilde{p})}{2} \quad (13)$$

クラスター間の距離の計算手法としてはウォード法¹⁸⁾を採用する。

4. 実験結果

本節では、提案手法を実際のエネルギー消費データに適用した結果から、提案手法により典型的なパターンが抽出できることを示す。

本研究では国際的なプロジェクト“ Annex 42, The Simulation of Building-Integrated Fuel Cell and Other Cogeneration Systems ”のデータ“ Generated Canadian Occupant-driven Electrical Load Profiles(EXCEL format, low electricity consumption) ”を用いて解析を行った¹⁹⁾。このデータはカナダのある家庭の電気消費量のプロファイルである。デー

タは約 3 年分、5 分間刻みで電力消費量をモニタしたものであり、1 日のデータは 288 次元のベクトルで表すことができる。図 1 は対象となる家庭のある 2 日間の電気消費量 (単位

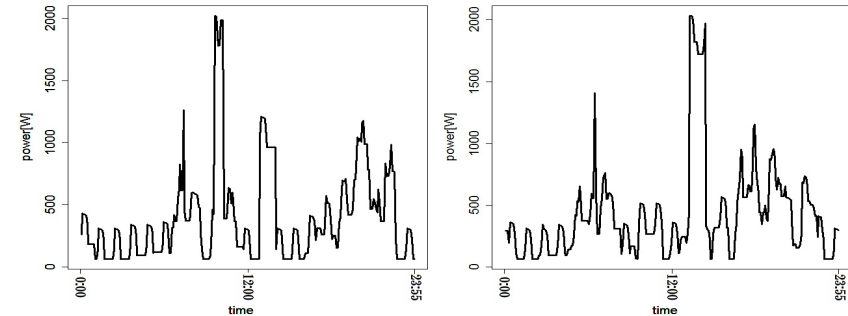


図 1 Examples of energy consumption data of a household in a day.

は W) のデータである。この 2 つのプロットから、1 日に電気消費のピークが複数現れることが分かる。一方、これらのピークの間隔や大きさは異なっていることも分かる。こうしたピークの間隔と大きさを混合正規分布でモデル化し、全データを階層的クラスタリング手法を用いていくつかの典型的なパターンに分類する。

家庭には、その家庭の生活スタイルがあり、その決まった生活スタイルにより電気消費パターンも存在すると仮定する。図 1 に示したように、1 日に電気消費のピークは複数ある。このようなピークは混合正規分布で近似できると仮定し、混合正規分布のパラメータで 1 日の電気消費量を近似する。混合正規分布のパラメータは EM アルゴリズムで計算する。本実験では、R 言語²⁰⁾ の“ mclust ”というパッケージの中の“ Mclust ”という関数を用いる。混合正規分布の分布数を任意に設定できるが、分布数を増やすとモデルが複雑になる。解析を簡単にし、オーバーフィットを防ぐために、パラメータの数を抑えた方がよい。モデルを簡単化するために、消費電力のピークは朝、昼食から午後、夕食から夜の各々 3 つの時間帯で現れると仮定する。その上で、混合正規分布の混合数を BIC 情報量基準²¹⁾ で 1 から 3 まで最適な混合数を推定する。

次に、推定した混合正規分布のパラメータを用いて、(13) より全ての 2 日間の組について距離を計算する。計算された距離を用いて、階層的クラスタリングによってその家庭の電気消費量を典型的なパターンに分類する。

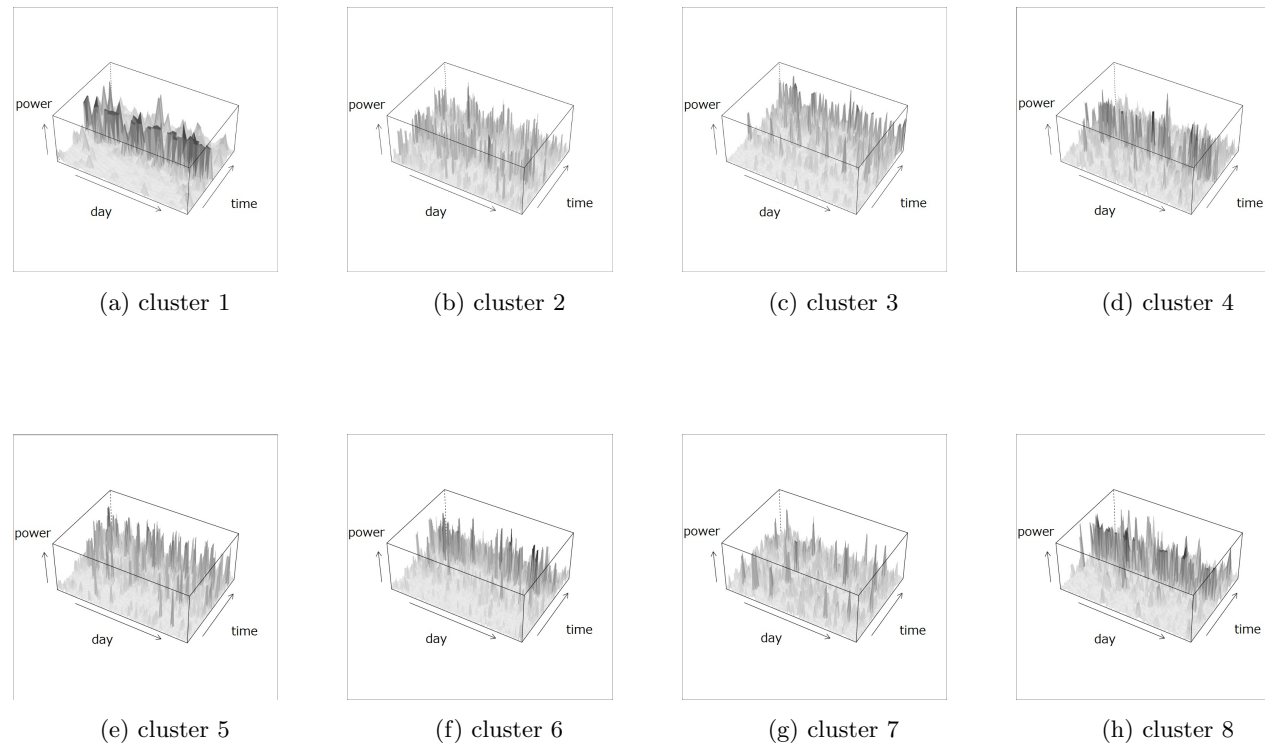


図 2 3-dimensional plots of energy consumption data within each cluster in a household.

図 2 に、実験で得られた 8 つのクラスタを示す。 (a) から (h) までそれぞれの図の縦軸は電力消費量 (単位 W) を表し、横軸は一日の時刻を表す。 図 3 はそれぞれのクラスタ内に含まれるデータの平均と標準偏差を表したものである。 図 2, 3 より、次のことが分かる:

- クラスタ 1, 4 と 8 は昼間前後に一つ大きなピークが出るクラスタである。
- クラスタ 2 は昼間前後と夕方に 2 つピークが出るパターンになっている。
- クラスタ 3, 5 と 6 は夕方から夜までの間に一つ大きなピークが出るクラスタである。
- そのほか、クラスタ 7 ははっきりしたピークが出ず、待機電力などによる定常的な電力消費に対応していると考えられる。

家庭用の発電システムの組み合わせを最適化するためには、それぞれの典型的なパターンが、全体の内どの程度の割合で生じているかも重要な情報である。 表 1 はそれぞれのクラスタに何日分のデータが入っているのかを表している。 この表より、この家庭の電気消費パターン

表 1 Number of elements in each cluster

Cluster	1	2	3	4	5	6	7	8
# of elements	36	289	100	114	133	145	106	80

は、クラスタ 2 が最も多く、クラスタ 5 と 6 も多く存在することが分かる。クラスタ 2 のピー

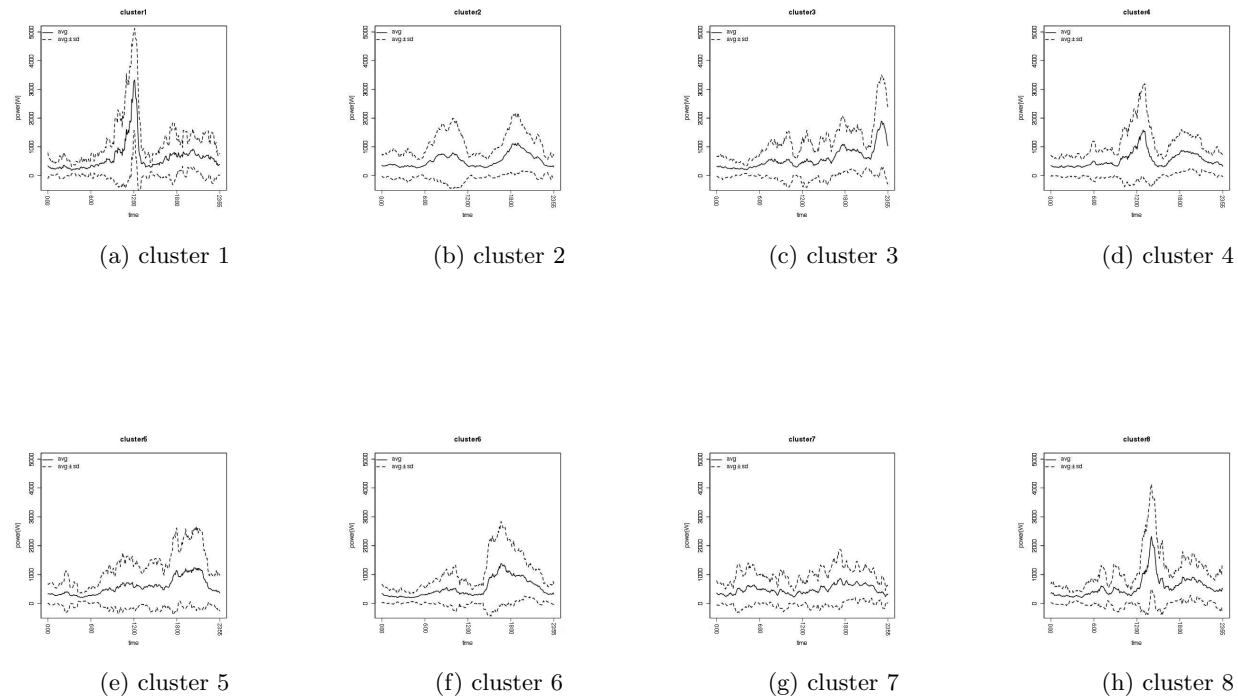


図 3 Average and one standard deviations of the basic energy consumption patterns in eight clusters extracted by the proposed method. Most of plots show clear peaks which characterizes the basic pattern.

クは昼頃と夜で、クラスタ 5, 6 は夜にピークが出ていることから、この家庭にとって最適な発電手段としては、夕方から夜にかけて多くの電力を供給できるものと推測できる。

5. 結 論

本研究では、家庭の電気消費を混合正規分布でモデル化し、典型的な消費パターンを抽出する手法を提案した。家庭内のエネルギー供給を最適化するために、それぞれの家庭の典型的な消費パターンの傾向を考える。電気消費データを推定されたパラメータで表し、8 クラ

スタにクラスタリングした。

家庭に発電システムを導入する際、初期投資、保守コストと導入後のメリットが重要な要素である。長い年月に渡り発電システムを利用する場合、その家庭の電気消費パターンが変化するという問題が生じる。例えば、その家庭に子どもが生まれたり、子どもが家から独立したりといった家族構成の変化により、典型的な電力消費パターンは大きく変化すると考えられる。家族構成以外にも、生活スタイルが変わると、電気消費パターンも変化してしまう。本研究では、*Annex42* の電気消費データについて解析したが、提案手法は任意の電気消

費データまたは電気供給データに適用できる。現在、特定の地区の多数の家庭での電気消費データが長期間に渡り収集されており、こうしたデータに本稿で提案する分析手法を適用することは家庭に導入する発電手段の組み合わせの最適化に有用であると考えられる。今後の重要な課題の一つは、太陽光発電、風力発電、コジェネレーションシステムなどの供給側発電パターンについて分析し、本研究で検討した需要側の解析と合せ、最適な発電システム導入を実現することである。

Acknowledgment

本研究の一部は、科研費（研究活動スタート支援）No.22800067 及び、私立大学研究高度化推進事業 No.S1001027 の補助を受けて行われた。

参考文献

- 1) A.Sayigh, "Worldwide progress in renewable energy," vol.3, no.4, 2008, pp. 86–90.
- 2) E.Martinot, A.Chaurey, D.Lew, J.R. Moreira, and N.Wamukonya, "Renewable energy markets in developing countries," *Annual Review of Energy and the Environment*, vol.27, no.1, pp. 309–348, 2002.
- 3) W.Grossmann, I.Grossmann, and W.K. Steininger, "Indicators to determine winning renewable energy technologies with an application to photovoltaics," *Environmental Science & Technology*, 2010.
- 4) C.Samaras, Ed., *Learning from Wind: A framework for effective low-carbon energy diffusion*, World Renewable Energy Congress. Elsevier, August 19-25 2006.
- 5) *10th International Conference on Intelligent Systems Design and Applications, ISDA 2010, November 29 - December 1, 2010, Cairo, Egypt.* IEEE, 2010.
- 6) Vringer, Kees, Theo Aalbers and Kornelis Blok, "Household energy requirement and value patterns," *Energy Policy*, 35 (1) p.p. 553-566, 2007.
- 7) M.L. Baughman, N.A. Eisner, and P.S. Merrill, "Optimizing combined cogeneration and thermal storage systems: An engineering economics approach," *Computer Engineering*, vol.4, no.3, 1989.
- 8) N.W. Alnaser, R.Flanagan, and W.E. Alnaser, "A comprehensive model for accelerating the building integrated photovoltaic (bipv) / wind turbine (biwt) construction projects in the kingdom of bahrain," *The Open Construction and Building Technology Journal*, vol.3, pp. 1–11, 2009.
- 9) A.B. Kanas-Patill, R.P. Saini, and M.P. Shama, "Sizing of integrated renewable energy system based on load profiles and reliability index for the state of uttarakhand in india," *Renewable Energy*, vol.36, pp. 2809–2821, 2011.

- 10) A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series. B*, vol.39, 1977.
- 11) C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- 12) T.M. Cover and J.A. Thomas, *Elements of information theory*. John Wiley and Sons, Inc., 1991.
- 13) J.R. Hershey and P.A. Olsen, "Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.*, vol.4, 2007, pp. IV–317–IV–320.
- 14) M.N. Do, "Fast approximation of Kullback-Leibler distance for dependence trees and hidden Markov models," *Signal Processing Letters, IEEE*, vol.10, no.4, pp. 115–118, March 2003.
- 15) A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.
- 16) A. Jain, M. Murty, and P. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol.31, no.3, pp. 264–323, 1999.
- 17) D. H. Johnson and S. Sinanovic, "Symmetrizing the kullback-leibler distance," Technical Report, Rice University., Tech. Rep., 2000.
- 18) J.H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol.58, no. 301, pp. 236–244, March 1963.
- 19) I. Beausoleil-Morrison, "An experimental and simulation-based investigation of the performance of small-scale fuel cell and combustion-based cogeneration devices serving residential buildings," 2008. [Online]. Available: http://www.ecbcs.org/docs/Annex_42_Final_Report.pdf
- 20) R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010.
- 21) G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol.6, no.2, pp. 461–464, 1978.