HOKKAIDO UNIVERSITY

Title	相互相関を考慮した非線形予測モデルに基づく 札幌市気温と北海道大学構内電力需要の同時推定
Author(s)	岩山, 浩将
Degree Grantor	北海道大学
Degree Name	学士(情報科学)
Issue Date	2012-03-22
Doc URL	https://hdl.handle.net/2115/52278
Type	bachelor thesis
File Information	Iwayama_BachelorThesis2011.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

相互相関を考慮した非線形予測モデルに基づく
札幌市気温と北海道大学構内電力需要の同時推定

岩山 浩将

北海道大学 工学部情報エレクトロニクス学科 情報工学コース 混沌系工学研究室

平成 23 年度 卒業論文

目 次

1 はじめに **1**
1.1 研究背景 1
1.2 本論文の構成 2

2 実データ解析からのエビデンス **2**
2.1 典型的 24 時間での推移と気温-電力需要曲線 2

3 時系列データの相互相関 **2**
3.1 ２体相互相関係数 3

4 予測モデル **4**
4.1 PUCK モデル 4
4.2 相互相関に基づく PUCK の改良 5
4.3 AIC によるパラメータ推定 5

5 数値実験 **6**
5.1 評価指標 6
5.2 同時推定 6

6 考察 **7**
6.1 比較 7
6.2 電力需要データの季節性について 7

7 おわりに **12**

A ポテンシャルの導出について **13**

1　はじめに

1.1　研究背景

近年, 複数の時系列データを用いた多変量解析の重要性が高まっている. このような複数の時系列を用いたデータ解析は様々な分野で使用されており, 計算機の性能向上や分散処理技術の発達により今後更に重要性が高まっていくと考えられる.

そこで, 我が国の問題に目を向けてみると, 2011 年 3 月 11 日に発生した東日本大震災以来, 各地の原発停止に伴い, 電力供給量が通年比で減少し, 空調設備の稼働率が高くなる夏期には, 電力需要が供給量限界に迫り, 各地の電力不足とその対策が逼迫した社会問題となったのは未だ記憶に新しい. 一方, あの原発事故以来, 我々は電力需要情報の提供側に対して「電力会社や政府行政が本当のことを言うか ?」という疑念を抱くようになったこともまた事実である. しかし, 電力会社や政府が真実を語ろうが語るまいが, 限りあるエネルギー資源を際限なく無駄使いしてよいことにはならない. 特に, 決して少なくない税金がその運営に投入される国や地方自治体の各種公共機関にとって, 常日頃から電力の無駄使いを控えることは義務ですらあるであろう.

事実, より局所的な電力需要に目を向ければ, 北海道大学札幌キャンパス内での電力需要に関し, (北海道電力との契約上) 設定された「上限使用電力」に対し, その何パーセントまで迫っているか, という情報が 30 分 (= 0.5 時間) 単位のデータとして北大ホームページ上で公開されるようになった [4]. 大学構内では, 空調設備はもちろんのこと, 日々の研究活動における実験機器の稼働, (学期

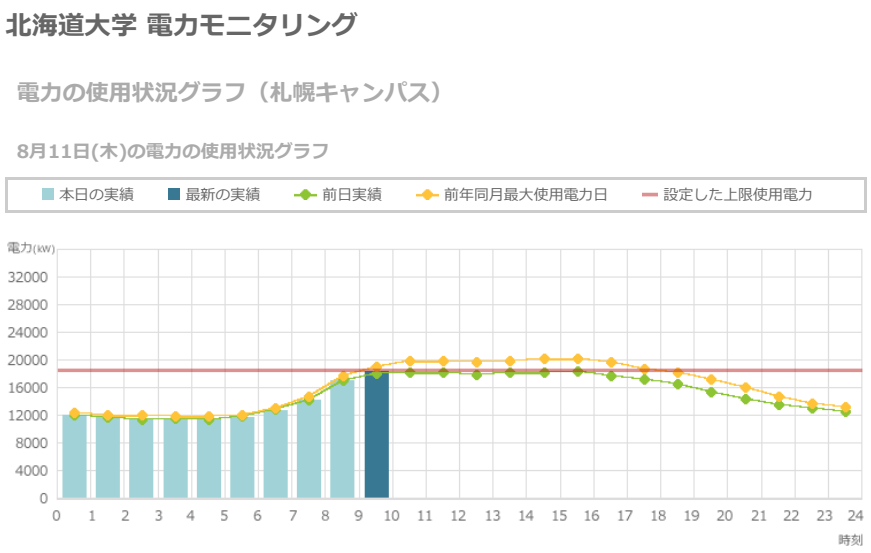

図 1: 昨年 8 月 11 日の北大構内の電力需用量の推移. 2012 年 2 月現在でも北大ホームページから閲覧できる.

間においては) 各講義室の照明, 計算機システム/サーバを維持するための電力等も総需要に対し大きな割合を占める. このとき, このような電力需要の時系列から電力利用状況を解析し, 可能ならば電力使用量を予測する方法を考えることは, 単に「面白い」という学術的興味に限らず, 実用的見地からも重要であろう. ところで, 北大の立地する札幌市の夏場は, 他の都府県の都市と比べ, 相対的に気温が低いことは周知の事実である. しかし, その札幌市民と言えど, 気温が上がればエアコンのスイッチを入れたくなるのもまた人情である. 結果, 北大構内の空調設備の稼働率が高くなることで電力需要が上昇することもまた容易に予想できる. 従って, これら「気温」「電力需要」両

者の「相関」は普遍的なものであり, 気温の上昇下降の時系列パターンを電力需要の予測に用いることは理にかなっている.

また, 電力需要データのような時系列を予測する手法は古くから多く提案されており,AR モデルやそのモデルを拡張した ARMA や GRACH モデル等の伝統的なモデルの他, 最近では物理学を基礎とした予測手法も提案されている [1],[2]. 中でも高安らは時系列の変動に移動平均を中心とした「ポテンシャル力」を仮定することにより,「バブル」や「クラッシュ」等の時系列の変動を説明するモデルを提案した [3]. このモデルでは, 急激な変動の周辺でポテンシャル関数の形状が変化することは説明されているが, 逐次的な予測や金融データ以外への適用とその考察は未だなされていない. そこで, 本論文では 2011 年 8 月ひと月の札幌市気温と北大構内電力需要時系列の相互相関を実データ解析し, そこで得られた知見に基づき, 気温と電力需要を同時推定する方法を提案し, その妥当性を統計的に評価する.

1.2　本論文の構成

本論文の構成は以下の通りである. 第 2 節では気温, 電力需要データの統計的性質や, 両者の相関について詳しく説明する. 第 3 節では複数時系列の相関を量る指標として相互相関係数を定義し, その相関を可視化する. 第 4 節では時系列解析のモデルとして高安らが提唱する PUCK モデルについて紹介を行い, 相関を用いて改良した提案手法を説明する. 第 5 節では, 実際の電力需要データに対してここまでのモデルでの数値実験を行う. 第 6 節では第 5 節での数値実験の結果を概観し考察を行う. 第 7 節では今後の課題等を述べる.

2　実データ解析からのエビデンス

2.1　典型的 24 時間での推移と気温-電力需要曲線

まずは北大構内の電力需要系列 D_t と札幌市の気温系列 T_t に関し, 夏場における「典型的な一日」(2011 年 8 月 1 日) を選んでプロットし, 結果を図 2 に載せる.

この両者をみると, 明らかに気温と電力需要のデータのおおまかな変化のパターンは互いに似ていることがわかる. そこで, D_t および T_t のセットを 2 次元平面上の点列 (D_t, T_t), $t = 1, \cdots, 48$ として, それぞれを結んだ曲線を図 3 に示す. この結果は所謂「TD チャート」[6] と同じ一般的な電力需要構造 (「変形 8 の字型」) を持つことがわかる. すなわち, 気温の上昇する朝から日中にかけて電力需要も上昇するが, 気温が下がり始める夕刻から夜にかけては, 気温の低下するスピードと比べて需要 ($\simeq$ 人々の活動量) はなかなか下がらず, 結果として図 3 のような「変形 8 の字」を描くことになる.

従って, 本研究は「札幌市」という我が国の夏期気温からみて, やや特異な都市, また「北海道大学構内」という特殊で, 限定されたコミュニティを扱っているとは言うものの, 気温-電力需要曲線の観点からは極めて一般的状況に対する時系列予測を扱っていると言え, また気温と電力需要には何らかの相関があることが見て取れる.

3　時系列データの相互相関

この節では前節で見られた電力需要量と気温の関係を定量化するための相互相関係数（ピアソン係数）を定義する.

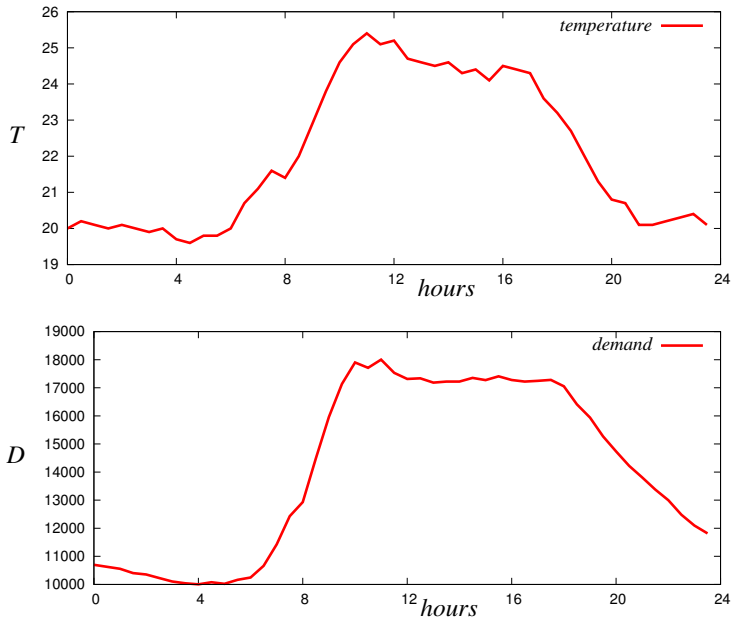

図 2: 2011 年 8 月 1 日の札幌市気温の推移 (上図) と北大構内電力需要の推移 (下図). 単位はそれぞれ「摂氏」「キロワット」.

3.1　２体相互相関係数

電力需要量と気温との相互関係を定量化するため, 相互相関係数 (ピアソン係数) を定義することにする. そこで, 気温, 電力需要データ系列の時刻 t での値 T_t, D_t に対し, 連続する 2 時刻間におけるこれらの値の変化量を対数尺度で測って $\Delta T_t = \log T_t - \log T_{t-1}, \Delta D_t = \log D_t - \log D_{t-1}$ で定義する. また, これらの量の過去 M ステップに渡る移動平均, 二体相互相関をそれぞれ次式で算出する.

$$\overline{\Delta T_t} \equiv \frac{1}{M}\sum_{l=t-M+1}^{t} \Delta T_l \tag{1}$$

$$\overline{\Delta T_t \Delta D_t} \equiv \frac{1}{M}\sum_{l=t-M+1}^{t} \Delta T_l \Delta D_l \tag{2}$$

このとき, 気温-需要間の相関係数は次式で与えられる.

$$\rho_{TD}^{(t)} = \frac{\overline{\Delta T_t \Delta D_t} - (\overline{\Delta T_t})(\overline{\Delta D_t})}{\sqrt{[\overline{(\Delta T_t)^2} - (\overline{\Delta T_t})^2][\overline{(\Delta D_t)^2} - (\overline{\Delta D_t})^2]}} \tag{3}$$

また,$\rho_{TD}^{(t)}$ は各時刻において,$-1 \leq \rho_{TD}^{(t)} \leq 1$ となる. このように定義した相互相関係数について, 図 4 に相関係数 ρ_{TD} と需要差分 ΔD の関係を時刻毎にプロットした結果を載せる. この図より, 相関係数はゼロではなく, 1 日の時間帯ごとに偏った値をとることがわかる. 例えば, 午前 4 時から 7 時の時間帯では両者は強い正の相関を持ち, しかも, この時間帯の縦軸は $\Delta D > 0$ であるから, この間, 気温と電力需要の双方が上昇することが容易に見てとれる. 同様にして, 午後 4 時から 7 時の時間帯では強い正の相関かつ $\Delta D < 0$ であるので, この間, 気温と電力需要双方が減少することがわ

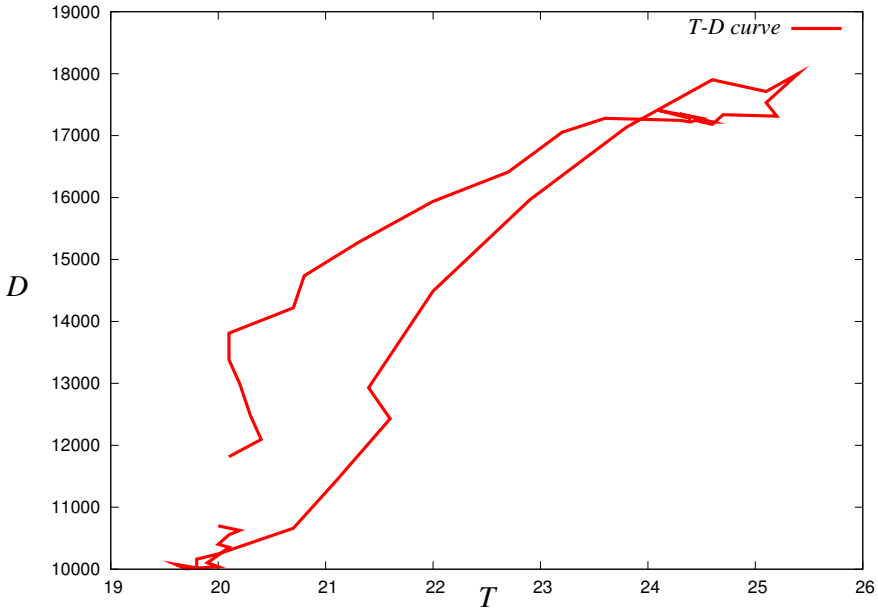

図 3: 気温 (T) と電力需要 (D) の関係. 所謂「変形 8 の字型」となることが見てとれる.

かる. 従って, 適切にこれら相関を取り込むことで, 気温, 電力需要を個別に予測するよりも, 推定精度の高い方法が構築できると期待される.

4　予測モデル

4.1　PUCK モデル

PUCK(Potential of Unbalanced Complex Kinetics) モデルとは株価のバブル, クラッシュを移動平均とポテンシャル力から推定するマクロな非線形時系列解析モデルである.
気温時系列データ T_t について,PUCK モデルは次式で定義される.

$$T_{t+1} = T_t - \frac{d}{dT}U(T,t)|_{T=T_t-T_t^{(M)}} + f_t \tag{4}$$

$f(t)$ はランダムノイズとしてのガウス分布で, 過去 M ステップにおける移動平均 $T_t^{(M)}$ は次式で定義される.

$$T_t^{(M)} = \frac{1}{M}\sum_{l=0}^{M-1} T_{t-l} \tag{5}$$

古典力学の「運動方程式」から類推することができるように, 気温にはポテンシャル U の勾配方向に気温を変化させる「ある種の力」が作用すると考える. 例えば, ポテンシャル U が $T_t = T_t^{(M)}$ に最小値を持つ「2 次曲線」で与えられるならば, 気温 T_t はその移動平均値 $T_t^{(M)}$ に近づくように更新され,(4) 式の f_t 項の影響で, 気温はその移動平均値 $T_t^{(M)}$ の周辺でガウス揺らぎを伴った値として予測される. 一方, ポテンシャル U が「3 次曲線」になれば,$T_t = T_t^{(M)}$ は不安定となり, 気温が急激に変動する状況を再現できる.

ここでは, 上記考察を踏まえて, 2 次, 3 次の曲線を含む形でポテンシャル関数 $U(p,t)$ をパラメトリックに

$$U(T,t) = \sum_{k=1}^{K} \frac{b_k(t)}{k+1} T^{k+1} \tag{6}$$

で与える. また, ポテンシャルの導出については, 付録 A の方法を用いることで実際のデータに対するプロットが可能である. 本論文では, このような PUCK モデルを予測の基本モデルとし, 次節で相関を用いてこれを拡張する.

4.2　相互相関に基づく PUCK の改良

前節でみたように, 気温と電力需要時系列の間には無視できない時間相関が存在する. そこで, 相互相関係数を用いて (4) 式を拡張する. 具体的には, 過去 Q ステップに渡る相関係数を用いて, 次式で T_t,D_t の予測式を以下のように書き直す.

$$T_{t+1} = T_t - \frac{d}{dT} U(T,t)|_{T=T_t - T_t^{(M)}} + \sum_{q=0}^{Q-1} \lambda_q \rho_{TD}^{(t-q)} + f_t \tag{7}$$

$$D_{t+1} = D_t - \frac{d}{dD} U(D,t)|_{D=D_t - D_t^{(M)}} + \sum_{q=0}^{Q-1} \lambda_q \rho_{TD}^{(t-q)} + g_t \tag{8}$$

右辺の第三項は相関項で, 過去 Q ステップについての相互相関係数と係数 λ から構成される. ここで g_t は f_t と無相関の白色正規ノイズである.
ここで, ポテンシャル関数 $U(T,t)$,$U(D,t)$ は (6) 式同様に

$$U(D,t) = \sum_{k=1}^{K} \frac{b_k(t)}{k+1} D^{k+1} \tag{9}$$

で与えるものとする.

次節では, このような相関項付 PUCK モデル（以下, 改良 PUCK）を前節の PUCK と数値実験により比較する.

4.3　AIC によるパラメータ推定

ポテンシャル関数 (6)(9) のパラメータ数 K とその値の決定には, 次式の赤池情報量規準 (AIC) を用いる.

$$AIC_T = -2ln(l(b_1, b_\gamma, \gamma, M, \lambda)) + 2K \tag{10}$$

$$\tag{11}$$

$$AIC_D = -2ln(l(b_1, b_\gamma, \gamma, M, Q, \lambda)) + 2K \tag{12}$$

$$\tag{13}$$

本論文では, パラメータ Q を常に $Q=1$ で固定する. 従って, K がここで決定すべきパラメータ数であり, $l(\cdots)$ は $w[f_t]$,$w[g_t]$ を正規分布で与えた際の尤度関数である. これらは白色正規ノイズの性質から, 具体的に

$$l(b_1, b_\gamma, \gamma, M, \lambda) = \prod_{t=n+N-1}^{n} w[f(t)] \tag{14}$$

$$l(b_1, b_\gamma, \gamma, M, Q, \lambda) = \prod_{t=n+N-1}^{n} w[g(t)] \tag{15}$$

で与えられることになる.

5　数値実験

我々は前節で導入した提案モデルの妥当性を検証するため,「北海道大学サステイナブルキャンパス推進本部」より提供頂いた 2011 年 8 月の休日を除いた 23 日 分の電力量を最大値 18000 [キロワット] で規格化したデータ, および, 2011 年 8 月の最高気温で規格化した札幌市気温データに対して数値実験を行う.

5.1　評価指標

評価指標としては一日 24 時間における各時刻 t での予測値 T_t, D_t と真値 $\hat{T}_t, \hat{D}_t$ の二乗誤差を選び, これを 23 日分のデータで平均した

$$\varepsilon_1^{(T)}(t) = \mathbb{E}[(T_t - \hat{T}_t)^2] \tag{16}$$

$$\varepsilon_1^{(D)}(t) = \mathbb{E}[(D_t - \hat{D}_t)^2] \tag{17}$$

を性能評価指標として採用する. ここに, 記号 $\mathbb{E}[\cdots]$ は 23 日分のデータのばらつきに関する平均を意味する. また, この二乗誤差とは別に, 本研究では気温, および, 電力需要の上下変動の符号に関する誤差も評価する. つまり,

$$\mathcal{T}_t \equiv \mathrm{sgn}(T_{t+1} - T_t), \tag{18}$$

$$\hat{\mathcal{T}}_t \equiv \mathrm{sgn}(\hat{T}_{t+1} - \hat{T}_t), \tag{19}$$

$$\mathcal{D}_t \equiv \mathrm{sgn}(D_{t+1} - D_t), \tag{20}$$

$$\hat{\mathcal{D}}_t \equiv \mathrm{sgn}(\hat{D}_{t+1} - \hat{D}_t) \tag{21}$$

とすると誤差:

$$\varepsilon_2^{(T)}(t) = \frac{1}{2}(1 - \mathbb{E}[\mathcal{T}_t\hat{\mathcal{T}}_t]),\ \varepsilon_2^{(D)}(t) = \frac{1}{2}(1 - \mathbb{E}[\mathcal{D}_t\hat{\mathcal{D}}_t]) \tag{22}$$

を第 2 の評価指標として採用する. 次節では, これらの指標を用いて気温と電力需要の同時推定を行った.

5.2　同時推定

2011 年 8 月の北大構内の電力需要データと, その時期の札幌市内の気温データに対して時系列推定を行った. 休日分を除いた平日 23 日分のデータそれぞれに対して, 指標 $\varepsilon_1, \varepsilon_2$ を計算した結果を図 5,6 に, 気温-電力需要の推定結果を用いた T-D チャートを図 7 に示す.

6　考察

6.1　比較

今回の実験結果では, 推定結果の誤差棒が大きく, 既存の PUCK モデルとその相関を用いた改良版である提案手法の優劣に関して結論を出すことは難しい. しかし, 一つの傾向として, 指標 $\varepsilon_1, \varepsilon_2$ ともに, 気温と需要が急激に変化する午前中から日中にかけての時間帯や, 夕刻から夜間の間において, 提案手法が平均的にみてより良い精度を与えていることが見てとれる. このような結果になった原因のひとつとして, 電力需要データの低頻度性があげられる. 今回使用した 30 分毎の電力需要データは, 時系列予測でより一般的に扱う株価や為替というような取引ごとに変動するデータに比べて非常に低頻度なデータである. 本研究で, 相関を加えるモデルとして採用した PUCK モデルは高頻度な金融データに対して用いられており, 本研究のような低頻度データに対して適用することによる影響やポテンシャルの形状の変化は現時点で考察されていない. さらに, 今回のデータの期間は 2011 年 8 月 1 日から, 同月 31 日の 31 日間のうちの平日の 23 日分と非常に短い期間だったことも関係している可能性がある. これらについては, 今後データ数を増やす等で再検証が必要である. また次節で扱うように, 電力需要データには一日毎に強い周期性があり, そのような周期性がポテンシャルに与える影響を考察することは今後の課題である.

6.2　電力需要データの季節性について

本研究では取り扱わなかったが, 電力需要データには強い季節性があると考えられている. このようなデータに対しては, 時系列モデルを「トレンド+周期性+誤差」で表わした $t(i) + s(i) + f(i)$ の季節性モデルが提案されており [7], このトレンド成分に対して前述の手法による予測を行い, 周期性成分と組み合わせることによる時系列予測はより高精度の予測に役立つものと考える.

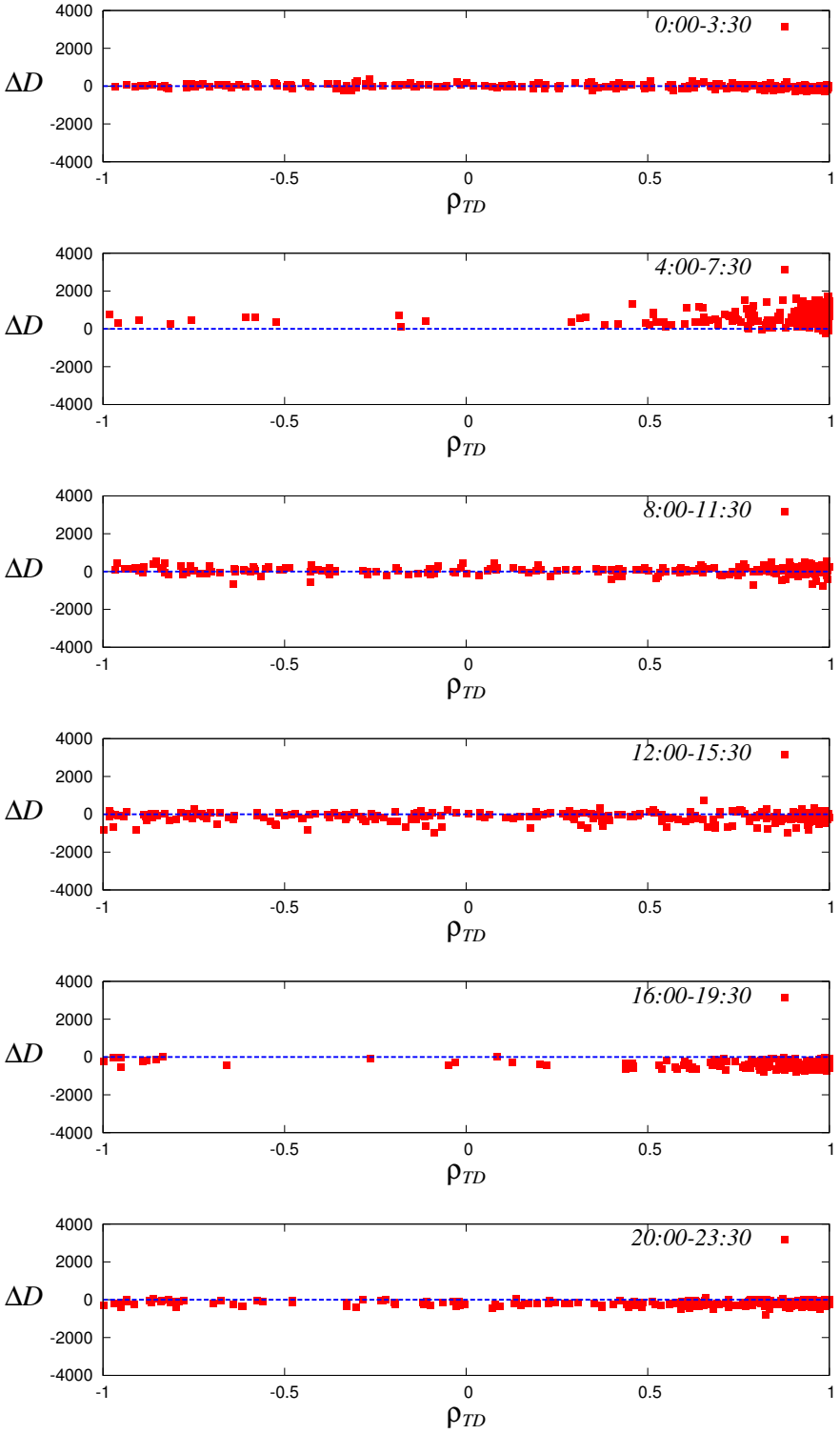

図 4: 相関係数 ρ_{TD} と需要差分 ΔD の関係.

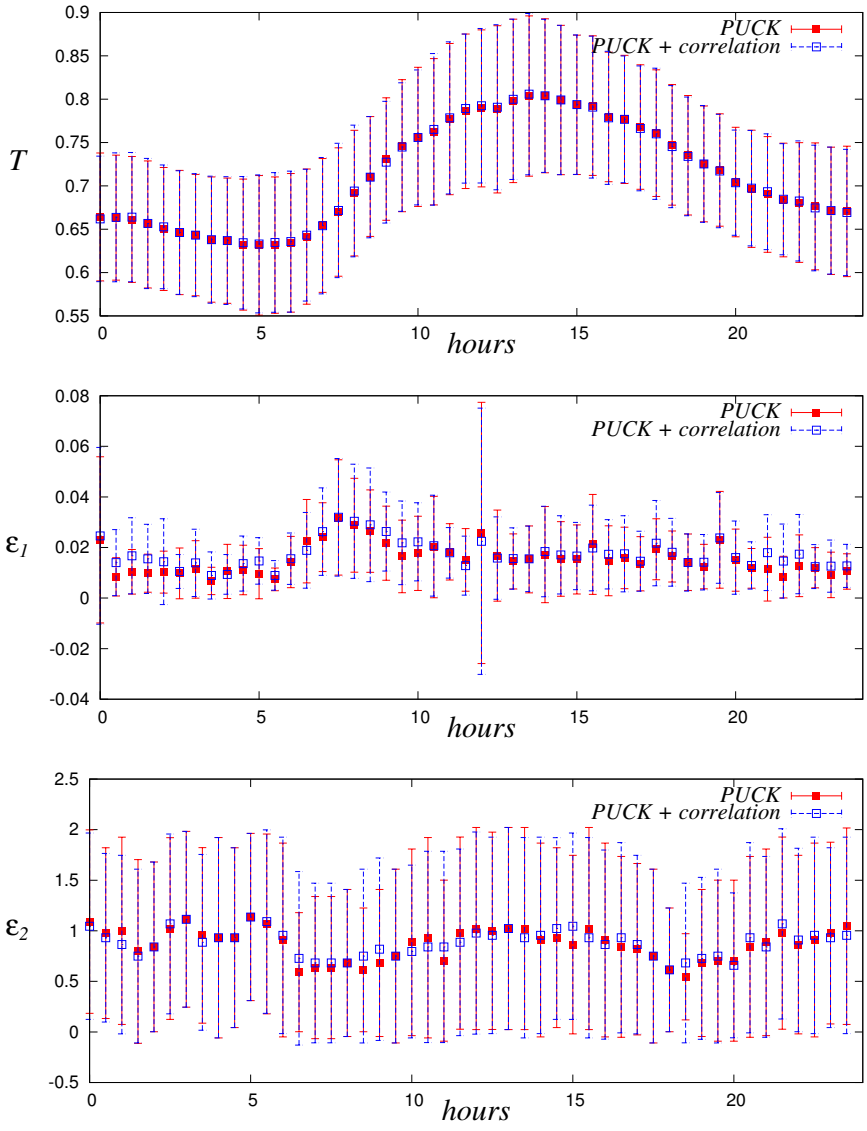

図 5: 上から札幌市気温の予測曲線, 誤差 $\varepsilon_1^{(T)}, \varepsilon_2^{(T)}$ の時刻依存性. 誤差棒は 23 日分のデータから算出された標準偏差.

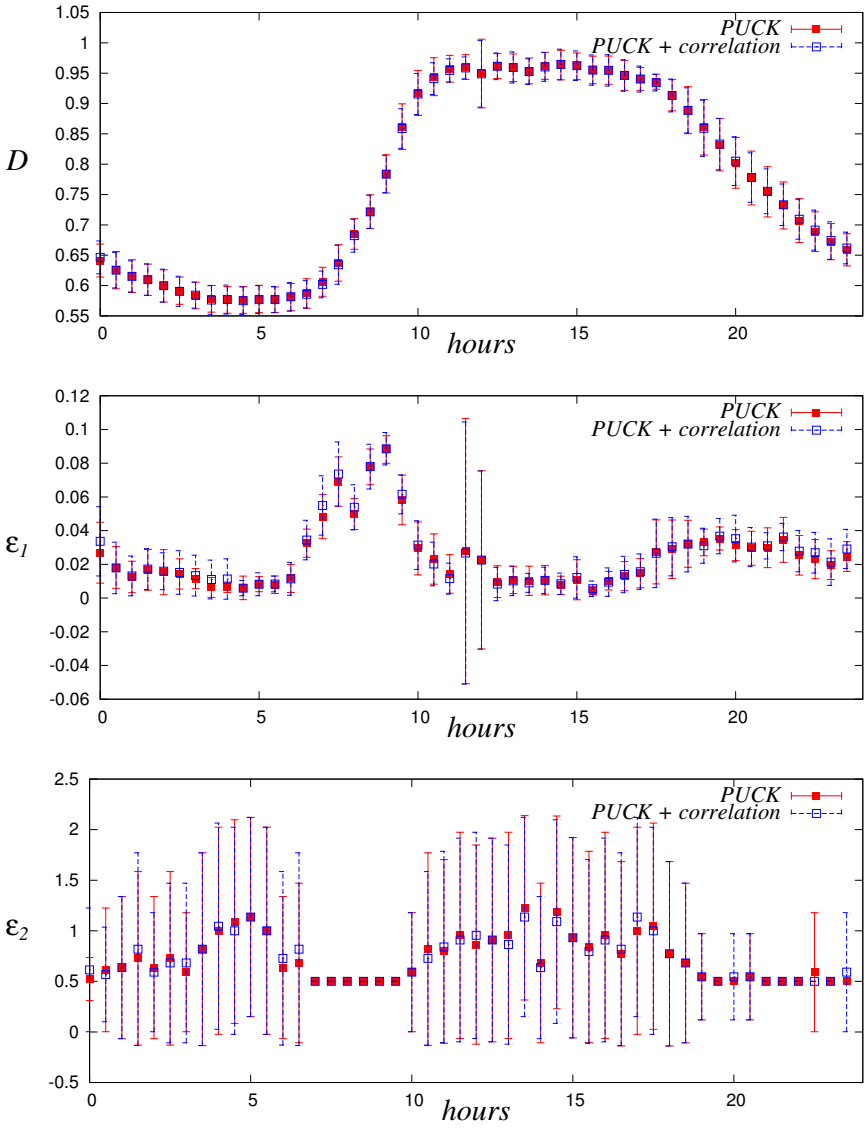

図 6: 上から電力需要の予測曲線, 誤差 $\varepsilon_1^{(D)}, \varepsilon_2^{(D)}$ の時刻依存性. 誤差棒は 23 日分のデータから算出された標準偏差.

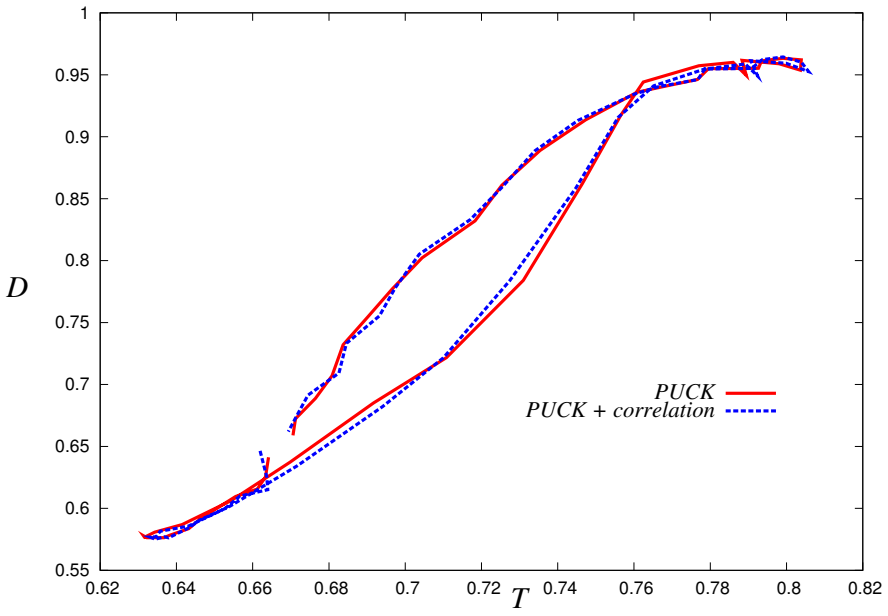

図 7: 気温 (T) と電力需要 (D) の関係. 所謂「変形 8 の字型」となることが見てとれる.

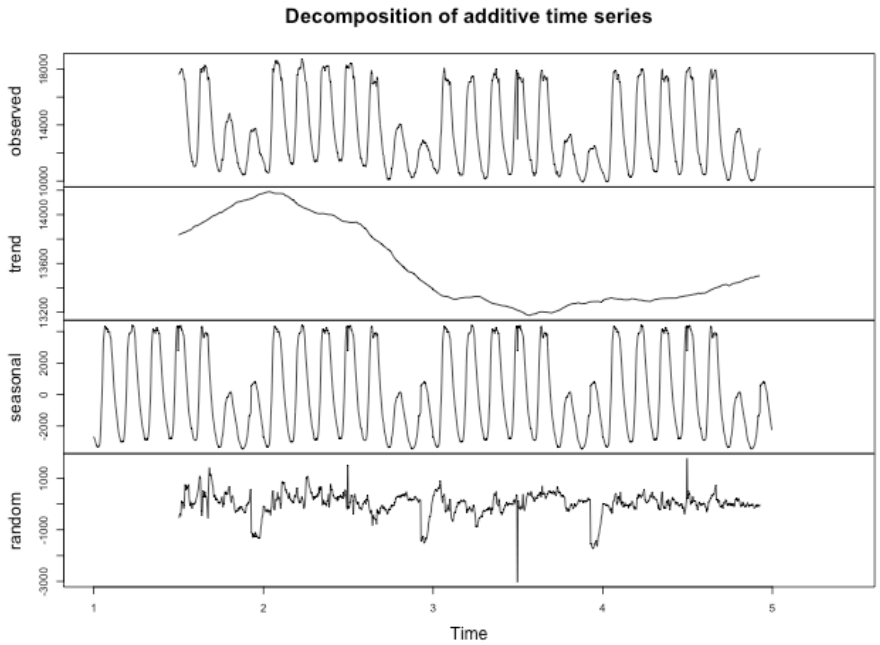

図 8: 季節性分割の例 (R を使用)

7　おわりに

本稿では札幌市気温と昨年 8 月の北大構内電力需要時系列の相互相関を実データ解析し, そこで得られた知見に基づき, 気温と電力需要を同時推定する方法を提案した. 本研究をスタートしたのは, 電力需給の問題が社会的に顕在化しはじめた 2011 年 6 月頃であり, 提供されたデータは 8 月のひと月分であった. 従って, より詳しい解析のためには, より多くのデータの提供を待たなければならない. 特に, 「春夏秋冬」「学期」等の季節性による変動を考慮することは重要である. また, ミクロなエージェントの行動モデルからマクロな需要時系列の構築 [8, 9], とりわけ, 頻繁に発せられる学内節電警報が需要変化に与える (時間差をともなった) 影響は興味深く, 今後の課題である. さらに, 電力需要においては, その (1 日における) 最大値の予測がより重要とされている. 例えば, 予め仮想的に設定された「限界レベル」を越えるまでの電力需要の時間発展は, ある種の確率過程である所謂 *First-passage process* で記述できる [10, 11, 12]. これらの知見を用いた予測法を検討することも興味深い課題と位置づけられる.

謝辞

指導教員である井上純一准教授に感謝します. また, 小野哲雄教授, 棟方渚助教をはじめとする, 研究室の皆様の協力にこの場を借りて感謝します. 最後に, 北大構内の電力需要データを提供していただき, 同システムについて丁寧かつ貴重な説明をして頂いた北海道大学サステイナブルキャンパス推進本部の森本智博氏, 横山隆氏に感謝します.

参考文献

[1] T. Kaizoji, Physica A 287, 493 (2000).

[2] J.-P Bouchaud and R.Cont, The European Physics Journal B 6, 543 (1998).

[3] W.Watanabe, H. Takayasu and M. Takayasu, Phys. Rev. E 80, 056110 (2009).

[4] http://www.facility.hokudai.ac.jp/denryoku/

[5] http://www.data.jma.go.jp/obd/stats/etrn/index.php

[6] 行木靖史, 日本オペレーションズ・リサーチ学会, 春季研究発表会, 1-A-7 (1995).

[7] 北川源四郎, 「時系列解析入門」, 岩波書店 (2005).

[8] 伊吹勇郎, 日向野隼輔, 井上純一「銘柄間相互相関に基づく金融危機の可視化と予測」人工知能学会研究会概要集 **SIG-FIN-008-04**, pp. 18-25 (2012).

[9] 日向野隼輔, 井上純一「平均場イジング模型の臨界現象と時系列予測」信学技報 **NC2011-108**, pp. 65-70 (2012).

[10] J. Inoue and N. Sazuka, *Physical Review E* **76**, 021111 (2007).

[11] N. Sazuka, J. Inoue and E. Scalas, *Physica A* **388**, pp. 2389-2853 (2009).

[12] J. Inoue and N. Sazuka, *Quantitative Finance* **10**, pp. 121-130 (2010).

A　ポテンシャルの導出について

まず, ポテンシャルとリターン $P(t+1)-P(t)$, 価格と移動平均の差 $P(t)-P_M(t)$ の関係:

$$\frac{P(t+1)-P(t)}{P(t)-P_M(t)} \propto -\frac{d}{dp}U(p,t) \tag{23}$$

が成り立つことに注意する. また, 恒等式:

$$\int_0^p -\frac{d}{dp}U(p,t)dp = -U(p,t)+U(0,t) \tag{24}$$

が満たされることに着目する. (23) 式を (24) 式左辺に代入すると, 積分変数を p から $P(t)-P_M(t)$ に変換して (積分区間も変更されることに注意)

$$\begin{aligned}
& -U(P(t)-P_M(t),t)+U(P(0)-P_M(0),t) \\
\equiv\ & -U(P(t)-P_M(t))+U(P(0)-P_M(0)) \\
=\ & \int_{P(0)-P_M(0)}^{P(t)-P_M(t)} \left\{\frac{P(l+1)-P(l)}{P(l)-P_M(l)}\right\} \Delta\{P(l)-P_M(l)\} \\
=\ & \sum_{l=0}^{t} \left\{\frac{P(l+1)-P(l)}{P(l)-P_M(l)}\right\} [P(l+1)-P(l)-\{P_M(l+1)-P_M(l)\}] \\
=\ & \sum_{l=0}^{t} \frac{(P(l+1)-P(l))^2}{P(l)-P_M(l)} - \sum_{l=0}^{t} \frac{(P(l+1)-P(l))(P_M(l+1)-P_M(l))}{P(l)-P_M(l)}
\end{aligned} \tag{25}$$

が得られる. ここに

$$d\{P(l)-P_M(l)\} = \Delta\{P(l)-P_M(l)\} = P(l+1)-P(l)-\{P_M(l+1)-P_M(l)\} \tag{26}$$

を用いたことに注意されたい. 従って, ポテンシャル $U(P(t)-P_M(t))$ は次式で与えられる.

$$\begin{aligned}
U(P(t)-P_M(t)) &= U(P(0)-P_M(0)) \\
&+ \sum_{l=0}^{t} \frac{(P(l+1)-P(l))(P_M(l+1)-P_M(l))}{P(l)-P_M(l)} \\
&- \sum_{l=0}^{t} \frac{(P(l+1)-P(l))^2}{P(l)-P_M(l)}
\end{aligned} \tag{27}$$

従って, ある時刻 t_* のときの「横軸」$P(t_*)-P_M(t_*)$ の値に対する「縦軸」(ポテンシャルの値) を計算するためには, (27) 式の積分値の上限を t_* として l に関する和を $l=0$ から $l=t_*$ まで計算していけばよい時刻 t_* を様々変えながら選んで得られる横軸 $P(t_*)-P_M(t_*)$ に対し, 上記を繰り返すことで「縦軸」であるポテンシャル値が数値的に得られていくことになり, これらの「点列」がポテンシャルの形状を与えることになる (時間原点 $l=0$ は適当に選択する).