

measure of credibility reflects the difficulty of the prediction solar energy supply.

## JIT モデリングによる太陽光発電量予測とその信頼性評価

沈 浩 洋<sup>†1</sup> 日 野 英 逸<sup>†1</sup>  
村 田 昇<sup>†1</sup> 若 尾 真 治<sup>†1</sup>

近年、エネルギー問題を解決するための新エネルギー技術が急速に発展している。その一つである太陽光発電システム (PV システム) は、クリーンで限りないエネルギーを利用できるという特徴から注目を集めており、実用化も進んでいる。一方、PV システムは天候に左右されやすく、出力が不安定という問題点を有する。急激な天気の変化により、PV システムによる発電量は短時間に大きく変動し、その予測は困難であることが知られている。本研究では、Just-in-time (JIT) モデリングにより PV システムの発電量予測モデルを構築し、その予測の信頼性尺度として Shannon 情報量を利用することを提案する。提案手法により、次の時刻の発電量予測値が信頼できるかどうかを事前に判断することが可能となり、急激な出力変化への対策を講じることが可能となる。本稿では、実験システムを用いて取得した太陽光発電データを用い、提案手法の有効性を検討した結果を報告する。

### JIT modeling of solar power prediction with a measure of credibility

Recently, remarkable developments of new energy technologies have been achieved against various energy problems. Photovoltaic (PV) system, one of such technologies, has an advantage of utilizing infinite and clean energy. On the contrary, it also has a disadvantage of unreliable power supply mainly caused by unstable weather. The fluctuation of the power supply of PV systems are considerably large because of rapid insulation changes and rapid weather changes, and in some cases, it seems impossible to realize high-accuracy prediction even with sophisticated prediction models. In this study, using a recently proposed estimator for the Shannon information content, a method to output a measure of credibility for prediction is proposed. With the proposed method, it is possible to judge whether the energy supply at a certain future time is unpredictably fluctuate compared to the current value or not, and it is possible to take measures against the rapid change of solar energy generation in advance. From an experimental result using solar energy supply data, we see that the proposed

#### 1. はじめに

太陽光発電システム (PV システム) は環境問題やエネルギー問題を解決するために有効な手段の一つとして期待されている。PV システムの運用管理における技術的な問題点の一つが、PV システムから商用系統への逆潮流である。逆潮流が生じると、電力システム全体に悪影響を及ぼし、経済的損失にもつながる。PV システムのより安定した運用のためには発電量の予測が不可欠である<sup>1)–3)</sup>。

発電量のような時系列データに対する予測と回帰手法は古くからさかんに研究されている。線型回帰<sup>4)</sup> は最もシンプルで頻繁に利用される手法である。線型回帰は、非線型関数を用いた変換により非線型回帰に拡張が可能である<sup>5)</sup>。また、線型回帰の回帰係数に何らかの制約を加える正則化回帰手法もさかんに研究されており、多くの実問題で成果をあげている。また、直接的に非線型モデルを用いて予測モデルを構築するアプローチも多く、例えば PV システムの発電の予測においてはニューラルネットワークの利用が検討されている<sup>2),6)</sup>。一方、機械学習の分野ではカーネル法による回帰手法が数多く開発されている。カーネルトリックを用いた代表的な回帰手法として、サポートベクターマシンによる回帰 (SVR) と関連ベクターマシンによる回帰 (RVM) がある<sup>7)</sup>。

非線型、非定常なデータに対する予測手法としては、局所回帰モデル (Just-In-Time, JIT) が有効である<sup>8)</sup>。JIT モデリングは非線型な予測モデルを簡単に実現できるため、発電量の予測にも用いられている<sup>9),10)</sup>。一方、太陽光発電を代表とする再生可能エネルギーは種々の変動要因の影響を強く受けってしまうことが多く、発電量の急激な変動を予測するのは、JIT モデリングを用いたとしても困難である。従って、発電量の実測値が計測される前に、予測値の信頼性について評価することが重要である。予測の信頼性を計算することができれば、システムに悪影響を与えそうな変動を事前に検知ことができ、必要な対策を講じることができる。

本研究では、予測値の信頼性の評価基準として Shannon 情報量を用いた手法を提案する。

<sup>†1</sup> 早稲田大学 先進理工学研究科 電気情報生命専攻

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

この基準は JIT モデリングにより局所モデルを作った後、モデルの予測値の Shannon 情報量を計算することで信頼性を評価するというものである。提案手法には以下の長所がある：

- (1) JIT モデリングによる局所モデルとして任意の予測モデルに適用することができる。
- (2) JIT モデリング、情報量計算は容易に実装可能であり、特に情報量の計算は非常に効率的に行うことが出来る。

本稿は以下のように構成される。第 2 節では時系列データに JIT フレームワークを適用する手法について説明する。第 3 節では Shannon 情報量の概念とその効率的な計算法を簡単に説明し、情報量を JIT モデリングで取得した近傍データセットから計算する方法を提案する。第 4 節では我々の太陽光発電実験システムから得られたデータを用いて、種々の回帰手法による予測値の比較と予測値に信頼度を付加した結果を示す。

## 2. JIT モデリングによる予測

本節では、JIT モデリングの概念を紹介し、時系列データの予測問題における JIT フレームワークを定式化する。

通常の推定問題では、データ全体を単一のモデルで表現する。しかし、非線型性及び非正常性が強いデータを単一のモデルで説明することは困難であり、サンプルの数が大きくなった場合は計算コストがかかる。JIT モデリングでは、過去のデータをデータベース (DB) に格納し、必要に応じて動的に DB からデータを抽出し、簡単な局所予測モデルを予測の直前に構築する。以下に、JIT モデリングによる回帰の具体的な手続きを示す。時系列の時刻  $t$  の値  $x_t$  を予測するのに、時刻  $(t-s)$  から時刻  $(t-1)$  までのデータを使い、部分ベクトルを作る。

$$x_{t-s:t-1} = (x_{t-s}, \dots, x_{t-2}, x_{t-1}). \quad (1)$$

JIT モデリングでは、過去のデータを格納した DB から部分ベクトル  $x_{t-s:t-1}$  に近いデータセットを取り出し、局所モデルを構築する。ここで、近傍を定義するために、データ間に距離基準を決め、DB から部分ベクトル  $x_{t'-s:t'-1}$  を探す。この  $x_{t'-s:t'-1}$  は、予め定義した距離で測ったとき、 $x_{t-s:t-1}$  に近いものである。本研究では、大規模な DB に対しても効率的に計算が可能である Manhattan 距離を使い

$$d(x_{t-s:t-1}, x_{t'-s:t'-1}) = \sum_{k=1}^s |x_{t-k} - x_{t'-k}| \quad (2)$$

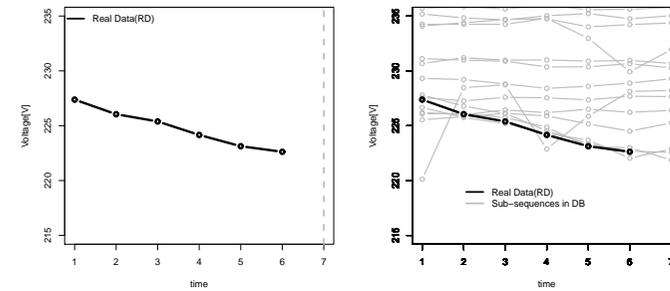
で部分ベクトル間の距離を計算する。次に、距離  $d(x_{t-s:t-1}, x_{t'-s:t'-1})$  に従い、DB から取り出した部分ベクトル  $x_{t-s:t-1}$  に最も近い  $M$  個の部分ベクトルの集合をデータセット  $D_t^p$

とする。 $D_t^p$  はデータセットを表すが、必要に応じて DB の添字集合を表すためにも同じ記号を利用する。データセット  $D_t^p$  を使い、 $x_t$  の予測モデルを構築する。JIT フレームワークによる局所予測モデルの選択は任意であるが、ここでは最も簡単な線型回帰モデルを使用するとして説明する：

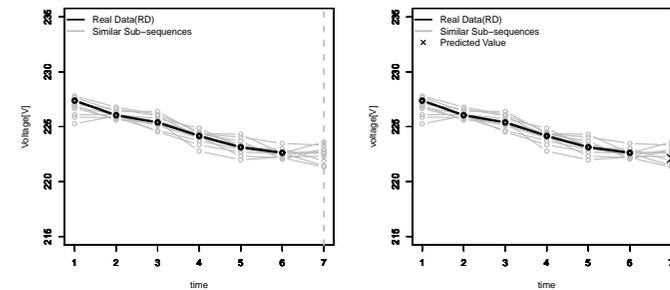
$$\hat{x}_t = \beta^\top x_{t-s:t-1} + \beta_0, \quad \beta = (\beta_1, \dots, \beta_s),$$

ここで、 $\beta_i, i = 0, \dots, s$  は予測値の二乗誤差が最も小さくなるように学習されたものである。

JIT モデリングの概要を図 1 に示す。



(a) a sub-sequence before the point to be predicted (b) sub-sequences in DB



(c) find M similar sub-sequences (d) predict the next point

図 1 Procedure of Just-In-Time modeling for time series prediction.

### 3. 情報量による信頼性指標

本節では、まず Shannon 情報量の概念について簡単に説明し、観測データから Shannon 情報量を計算する方法を簡単に示す。次に、JIT モデリングによって得られた予測値の信頼性の尺度として Shannon 情報量を使う手法を提案する。

#### 3.1 Shannon 情報量

ある注目している点 (検査点と呼ぶ) の Shannon 情報量は以下のように定義される

$$I_p(x) = -\log p(x). \quad (3)$$

ここで、 $p(x)$  は確率変数  $X$  の確率密度関数であり、 $x \in \mathbb{R}^d$  はその実現値である。 $E_p[\cdot]$  は密度関数  $p(x)$  に関する平均操作を表す。Shannon 情報量は情報理論<sup>12)</sup>、統計学<sup>4)</sup>、機械学習<sup>13)</sup> の分野で大きな役割を果たしている非常に基礎的な量である。また、Shannon 情報量  $I_p(x)$  の平均  $H(p) = E_p[I_p(X)] = -\int p(x) \log p(x) dx$  は、Shannon エントロピーとして知られている量である。

直観的には、式 (3) で定義されている Shannon 情報量は、観測データ  $x$  が確率密度関数  $p(x)$  からなるモデルから生成されていると仮定した場合の、観測データ  $x$  の「珍しさ」を表している。生じることが珍しいデータほどその予測が難しいと考えられるため、本研究では Shannon 情報量を予測の難しさの尺度として利用する。

#### 3.2 情報量の計算方法

観測データから Shannon 情報量を始めとする諸量を計算する問題は重要であり、多くの手法が提案されている。最もよく用いられる手法の一つとして、 $k$  近傍法 ( $k$ -NN) がある<sup>14)–16)</sup>。本研究では<sup>17),18)</sup> に基づき情報量を計算する。長さ  $s$  の部分ベクトル  $x_{t'-s:t'-1}$  が与えられた時、検査点  $\hat{x}_t$  の Shannon 情報量は次のように計算される

$$\begin{aligned} \Delta \hat{x}_t &= \hat{x}_t - x_{t-1}, \\ \Delta x_{t'} &= x_{t'} - x_{t'-1}, \\ \hat{I}(\hat{x}_t; x_{t'-s+1:t'}) &= \log 2 + 1 + \frac{1}{s-1} \sum_{x_i \in x_{t'-s+1:t'}} \log |\Delta \hat{x}_t - \Delta x_i|. \end{aligned} \quad (4)$$

ただし、 $x_i \in x_{t'-s+1:t'}$  は、 $x_{t'-s+1}$  から  $x_{t'}$  までの要素から 1 つずつ抜き出す操作を表す。このデータに基づく情報量の計算では時系列データが定常であることを前提としている。しかし、実験で扱う電圧のデータは定常とはみなせない。この問題を軽減するために、時系列の階差を取った後に Shannon 情報量を計算する。この情報量の計算はいくつかの望ましい性質を持っている (詳細は<sup>17)</sup> を参照)。中でも、従来の情報量推定よりも計算コストが少な

い点が重要である。多くの従来手法の計算量は  $O(s \log s)$  であるのに対し、式 (4) による情報量の推定に要する計算量は  $O(s)$  である。この性質から、全体プロセスを遅らせずに JIT モデリングの予測値の信頼性評価を行うことができる。

#### 3.3 予測の信頼性評価の手法

JIT モデリング過程により得られた近傍データセット  $x_{t'-s:t'-1} \in D_t^p$  を用いると、時刻  $t$  における予測値  $\hat{x}_t$  の情報量  $\hat{I}(\hat{x}_t; x_{t'-s+1:t'})$  を式 (4) を用いて求めることができる。さらに、DB から検査点  $x_{t'}$  より先の時刻に対応する、長さが  $s$  の部分ベクトル集合  $x_{t'+1:t'+s} \in D_t^f$  を用いて、 $x_t$  における予測値の信頼度を次式で定義する:

$$c(t, t') = \hat{I}(\hat{x}_t; x_{t'-s+1:t'}) + \hat{I}(\hat{x}_t; x_{t'+1:t'+s}). \quad (5)$$

$x_t$  の予測モデルを学習する前に、近傍データセット  $D_t^p$  と  $D_t^f$  の中の時刻  $t$  に対応する値  $c(t, t')$  を計算し、これらの (近傍集合における) 最大値を  $c(t)$

$$c(t) = \max_{t' \in D_t^p} c(t, t') \quad (6)$$

とする。もし、次の時刻の値の予測が困難な場合、 $c(t)$  が大きくなる。この性質を用いて、 $c(t)$  を JIT フレームワークで計算された予測値の信頼性評価基準とする。Shannon 情報量に基づく信頼性尺度の計算の過程を図 2 に示す。

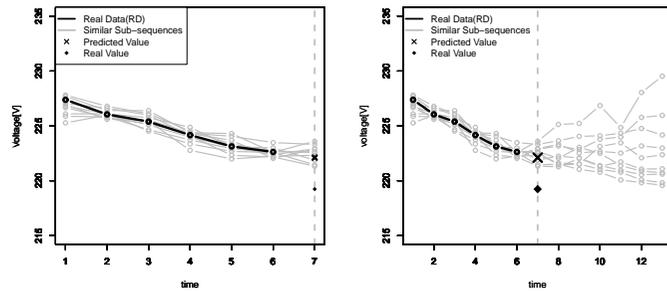
## 4. 実験結果

本節では、提案する JIT モデリングによる太陽光発電量の予測とその信頼性尺度の計算手法を、2005 年に早稲田大学の本庄キャンパスの PV システムから計測された 10 秒おきの電圧値データに適用した結果を示す。図 3 に発電量データの一例を示す。縦軸は電圧 (V)、横軸は時間である (日照時間帯として 6:30 から 16:30 までを切り出した)。図 3 から分かるように、晴天の場合は電圧値は滑らかに変動するが、曇りの場合には太陽が雲に隠れたり現れたりすることで電圧値の変動が激しくなる。

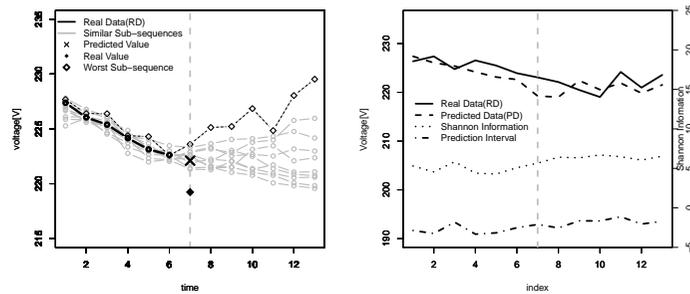
#### 4.1 電圧値予測の難しさ

種々の回帰手法により発電量を予測した結果を示す。

まず、線型回帰 (LR)<sup>5)</sup> により、一日の全時刻を単一のモデルでフィッティングした。次に、非線型回帰としてよく用いられる 3 層ニューラルネットワーク (NN)<sup>6)</sup> を適用した。また、 $\nu$ -SVR<sup>7)</sup>、 $\nu$ -RVM<sup>19)</sup> を使った予測を行った。次に、JIT モデリングにおいて選んだ近傍データを用いて線型回帰 (JIT+LR) 及び NN (JIT+NN) による局所予測モデルを学習する方法を適用した。最後に、時刻  $t$  の発電量  $x_t$  の予測値として 1 ステップ前の値  $x_{t-1}$  を用いた結



(a) M similar sub-sequences found in JIT procedure (b) find past and future sequences from DB



(c) estimated Shannon Information (d) result

図 2 Procedure of calculation of credibility measure for time series prediction.

果も考察する。

本稿で用いた回帰手法の多くはチューニングパラメタを有する。実験では、NNの隠れ層の数を30とし、SVRとRVMのパラメタ $\nu = 0.3$ 、JITモデリングにおいて選択する近傍のサイズを50個で実験を行った。また、時系列データを切り出すことでベクトルデータとして扱う際のベクトルの長さも適切に定める必要がある。今回の実験では、部分ベクトルの長さについて3パターンを試した。それぞれ、検査点から1, 3, 5分前のデータを用いた。なお、全ての実験はR言語 (version 2.9.1)<sup>20</sup> を使い実験を行った。

全ての実験において、予測値を評価するために2つの指標を採用する。 $T$ を評価対象とな

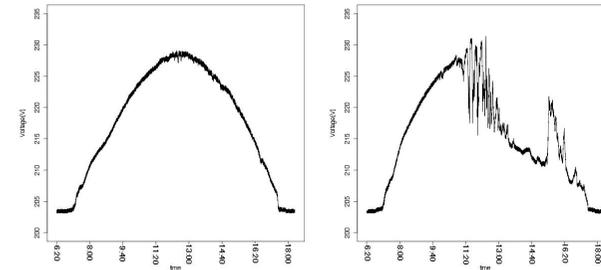


図 3 Examples of voltage profile of photovoltaic generation system at Honjo-Campus, Waseda University. Left: fine day where the change of voltage is smooth. Right: cloudy day where the change of voltage is unpredictable.

るサンプル点の数として、平均絶対値誤差 (AE) を

$$AE = \frac{1}{T} \sum_{t=1}^T |x_t - \hat{x}_t| \quad (7)$$

で定義する。一方、PVシステムによる発電量が一定のしきい値を超えると、PVシステムから商用系統へ逆潮流が生じる。本研究ではその閾値を222[V]と設定し、予測の評価尺度として、真の値がしきい値を超えているときに予測値もしきい値以上である、あるいは真の値がしきい値未満であるときに予測値もしきい値未満である場合を「的中」として、全データ点に対する的中した点の割合を的中率と定義する。表1に、種々の回帰手法による予測結果をまとめる。表1より、PVシステムの電圧の予測問題において、JITモデリングに基づ

表 1 代表的な回帰法による予測結果。AEとHRはそれぞれ、絶対値誤差と的中率の5日分データの平均を表す。括弧の中は5日分の標準偏差を表す。

	$s = 6$		$s = 18$		$s = 30$	
	AE	HR	AE	HR	AE	HR
LR	0.626(0.251)	0.977(0.015)	0.736(0.321)	0.972(0.018)	0.807(0.366)	0.969(0.020)
NN	5.052(1.933)	0.752(0.111)	5.193(2.001)	0.745(0.114)	5.216(2.025)	0.743(0.115)
RVM	15.190(9.559)	0.912(0.049)	23.491(15.433)	0.881(0.077)	27.053(17.767)	0.872(0.092)
SVR	0.835(0.354)	0.968(0.022)	1.041(0.503)	0.962(0.025)	1.148(0.553)	0.957(0.026)
JIT + LR	0.602(0.241)	0.977(0.015)	0.824(0.362)	0.970(0.018)	1.346(0.612)	0.956(0.029)
JIT + NN	0.601(0.253)	0.978(0.013)	1.172(0.422)	0.952(0.028)	1.167(0.422)	0.955(0.026)
Prev. Ref.	0.547(0.197)	0.979(0.013)	0.547(0.197)	0.979(0.013)	0.547(0.197)	0.979(0.013)

く手法が優れた予測性能を有することが分かるが、10秒前のデータをそのまま予測値として使った場合でも、他の手法と比べ高い精度で予測が出来ていることが分かる。これは、太陽光発電システムの発電量予測問題が非常に難しい問題であり、変動が激しい領域では回帰手法により得られた予測値よりも1ステップ前のデータをそのまま予測値として使ったほうが誤差は少ないということである。この結果は、予測値の値そのものよりも、その予測値の信頼性を評価することが重要であることを示唆している。

#### 4.2 予測値の信頼性の計算

本節では、太陽光発電システムの予測問題において Shannon 情報量を用いた信頼性の尺度を計算した結果を示す。

図4は、JITモデリングにより近傍の局所モデルを作り上げた後に計算された信頼性尺度のグラフである。変動が激しい時間帯では、信頼性尺度の値も大きくなっていることが分かる。つまり、予測値が信頼できないということになる。図5は変動が激しい一部分の拡大図である。図から分かるように、絶対値誤差(AE)が急に大きくなる時刻において、信頼性尺度の値も急に上がることが確認できる。特に、一番右の図を見ると、点線で引いた3つの時刻において、変動が大きいのが、その変動は滑らかで予測区間<sup>5)</sup>が狭くなっている。つまり、予測は容易であると見なされている。しかし、実際には絶対値誤差(AE)が大きく、その予測値は信頼できないと考えられる。提案手法による信頼性評価基準は、それらの時刻において大きな値を示し、予測値の信頼性が低いことを意味している。

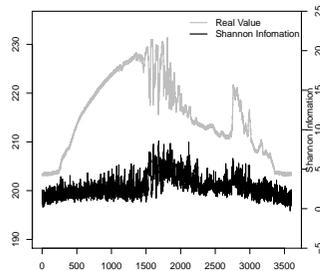


図4 The estimated credibility measures are high in the region that voltages are highly fluctuated.

## 5. 結 論

本研究では、従来の電力システムに自然エネルギーを導入するための重要な課題である太陽光発電システムの発電量の予測問題を扱った。太陽光発電のデータは非定常的な時系列データであり、パラメトリックな予測モデルによる予測は困難である。こうした非定常時系列の予測には、過去に蓄積されたデータを用いた Just-In-Time モデリングが有効であることが知られている。一方、急激な変動が頻繁に起こる発電量の予測問題は、JIT モデリングを用いても困難である。そこで我々は、JIT モデリングの予測結果に Shannon 情報量を信頼性の尺度として導入することを提案した。提案手法により、複雑な予測モデルや事後分布を使わずに過去のデータだけから予測値が信頼できるかどうかを評価できる。この手法は特に予測が難しい領域で大きな役割を果たす。

JIT モデリングと線型回帰を組み合わせた場合、予測の信頼性を測る尺度として予測区間を計算することが出来る。一方、予測区間の計算は簡単な線型モデルに対しては理論的に容易に計算できるが、一般の回帰モデルに対しては導出が困難である。予測精度を上げるために、JIT モデリングと複雑な予測モデルを組み合わせた時にも、提案する Shannon 情報量に基づく信頼度の評価は適用可能である。本研究では太陽光発電の予測問題を扱ったが、この信頼性を計算するフレームワークは他の時系列の予測問題にも適用可能である。今後の展開として、例えば、家庭のエネルギー消費量の予測に適用し、本稿で扱った発電量予測と合わせることで、電力の需要・供給モデルを構築することが考えられる。

JIT モデリングには莫大な DB からデータ間の距離を計算し、近傍データを抽出する必要があるため、計算量が大きいという問題点がある。本稿で扱う問題においては実用的な時間で近傍の計算ができたが、より大規模な DB を用いた予測のために、精度を落とすことなく DB を圧縮する手法は今後の重要な研究課題である。

## 謝 辞

本研究の一部は、科研費(研究活動スタート支援) No.22800067 及び科研費(基盤 B) No.23360128 の支援を受けて行われた。

## 参 考 文 献

- 1) G.S. Kinsey, K.Stone, J.Brown, and V.Garboushian, "Energy prediction of amonix cpv solar power plants," *Progress in Photovoltaics: Research and Applications*, pp.

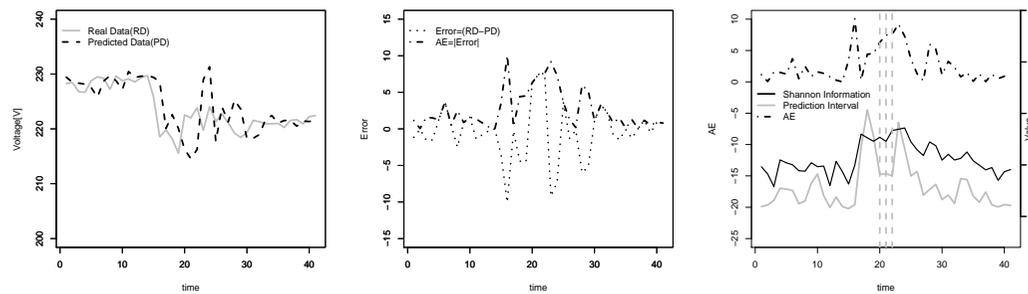


図 5 Highly fluctuated region are splitted from whole day data.

- n/a-n/a, 2010.
- 2) M.Kawaguchi, S.Ichikawa, M.Okuno, T.Jimbo, and N.Ishii, "Prediction of electric power generation of solar cell using the neural network," in *KES (2)*, 2006, pp. 387–392.
  - 3) T.-C. Yu and H.-T. Chang, "The forecast of the electrical energy generated by photovoltaic systems using neural network method," in *Electric Information and Control Engineering (ICEICE), 2011 International Conference on*, april 2011, pp. 2758 –2761.
  - 4) T.Hastie, R.Tibshirani, and J.Friedman, *The Elements of Statistical Learning — Data Mining, Inference, and Prediction*, Springer, first edition, 2001.
  - 5) A.J. Dobson, *An introduction to generalized linear models / Annette J. Dobson*, Chapman & Hall/CRC, Boca Raton :, 2nd ed. edition, 2002.
  - 6) B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, January 1996.
  - 7) B.Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, 2002.
  - 8) G.Cybenko, "Just-in-Time Learning and Estimation," in *Identification, Adaptation, Learning. The Science of Learning Models from data*, S.Bittanti and G.Picci, Eds., NATO ASI Series, pp. 423–434. Springer, 1996.
  - 9) D.Hosogoshi S.Wakao, "Fundamental study on the application of just-in-time modeling to wind power estimation," in *European Wind Energy Conference, 2006.*, 2006.
  - 10) H.Suzuki, Y.Watanabe, and S.Wakao, "Short-term pv output forecast using just-in-time modeling," in *European Photovoltaic Solar Energy Conference, Valencia, 2008.*, 2008.
  - 11) C.E.Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423,623–656, 1948.
  - 12) T.M. Cover and J.A. Thomas, *Elements of information theory*, John Wiley and Sons, Inc., 1991.
  - 13) C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
  - 14) L.F. Kozachenko and N.N. Leonenko, "Sample estimate of entropy of a random vector," *Problems of Information Transmission*, vol. 23, pp. 95–101, 1987.
  - 15) J.Beirlant, E.J. Dudewicz, L.Györfi, and E.C. Meulen, "Nonparametric entropy estimation: An overview," *International Journal of the Mathematical Statistics Sciences*, vol. 6, pp. 17–39, 1997.
  - 16) L.Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, pp. 1191–1253, June 2003.
  - 17) H.Hino and N.Murata, "A computationally efficient information estimator for weighted data," in *ICANN (2)*, 2011, pp. 301–308.
  - 18) 日野英逸 ほか：分位点に基づく重み付きデータの情報量推定手法とその応用, 第三回 IBISML 研究会 (2010).
  - 19) J.Shawe-Taylor and N.Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, New York, NY, USA, 2004.
  - 20) R.Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010.