

## 変化点検出法による農業環境データの分析 Analysis of Agricultural Environment Data based on Change Point Detection

小野 拓海<sup>†</sup>  
Takumi Ono

岩崎 清斗<sup>‡</sup>  
Kiyoto Iwasaki

大久保 誠也<sup>†</sup>  
Seiya Okubo

斉藤 和巳<sup>†</sup>  
Kazumi Saito

### 1. はじめに

株価や Twitter, レビューデータなどの, 多様な時系列データにおける変化点検出は, 社会状況の変化の察知や, ユーザの異常行動の把握など, 多くの用途で利用されている. 変化点検出手法の一つとして, ユーザの活動間隔のモデルを導入し, 尤度比検定を土台として再帰的に変化点を求める手法が提案されており [1], 株価の動きから経済活動の動きや影響力のある社会現象の抽出, twitter のバースト期間抽出によるネット上のイベント, 炎上問題の発見などの研究に応用されている. これらの手法は, 系列データの値が大きく変化した時刻を変化点とすることで, 何らかのイベントや変化要因の発見することができる. 他にも, ある既知のイベント時刻が変化点と見なされていない場合, そのイベントの影響度は低いと考えられるため, イベントの影響力の研究にも活用できる.

この手法を基に, ガウス分布でのモデリングの下で最小記述長原理を土台に逐次的に変化点を求める手法が提案されている [2]. 文献 [2] の手法はある区間を時系列データの平均値で近似したときの L2 誤差 (自乗誤差) が最小となるような時刻を変化点とし再帰的に求める (以下, この方法を L2cpd (change point detection) 法と呼ぶ). しかしながら, L2cpd 法は L2 誤差の最小化問題であるために, 外れ値による影響を受けやすいことが予測される. これに対し, 変化点間の時系列データの平均値ではなく, メディアン (中央値) を用いて L1 誤差 (絶対値誤差) が最小となるような変化点を検出する手法が提案されている [3] (以下, この方法を L1cpd 法と呼ぶ). 本稿では, 農業において作物の生育に関係の深い気温や湿度などの気象データ (農業環境データ) を用いて, これら手法で検出する変化点を比較し, 両者の違いを評価する.

### 2. 分析法

L2cpd 法は平均値の L2 誤差に注目して変化点を求める. 時系列データの時刻  $t$  の値を  $x_t$  とし, 時刻 1 から  $T$  までの時系列データを  $\mathbf{x} = (x_1, \dots, x_T)$  と表す. また, 変化点の個数を  $K$  とし, それぞれの変化点を古い順に  $F_1$  から  $F_K$  に格納する変化点時刻ベクトルを  $\mathbf{F}$  とする. 便宜上  $F_0 = 0$  かつ  $F_{K+1} = T$  と設定すると, ベクトル  $\mathbf{F}$  は  $(K+2)$ -次元ベクトルで表わされる. 区間  $F_{(k-1)} < t \leq F_k$  内の平均値を  $\mu_k$  と表すと, L2 誤差  $E(\mathbf{F})$  は次式で計算することができる.

$$E(\mathbf{F}) = \sum_{k=1}^{K+1} \sum_{t=F_{(k-1)+1}}^{F_k} (x_t - \mu_k)^2 \quad (1)$$

よって, 変化点検出問題は  $E(\mathbf{F})$  を最小化する  $\mathbf{F}$  を求める問題として定式化できる. 変化点時刻ベクトル  $\mathbf{F}$  を求める L2cpd 法の詳細は文献 [2] を参照されたい.

平均値ではなくメディアンによる L1 誤差に注目するのが L1cpd 法である. ここでメディアンによる L1 誤差を最小にする変化点時刻ベクトルを  $\mathbf{D}$  とする.  $\mathbf{F}$  同様,  $D_0 = 0$  かつ  $D_{K+1} = T$  と設定し, それぞれの変化点を古い順に  $D_1$  から  $D_K$  に格納する  $(K+2)$ -次元ベクトルとして  $\mathbf{D}$  を表す. 区間  $D_{(k-1)} < t \leq D_k$  の中央値を  $m_k$  とすると L1 誤差は以下の式で定義される.

$$G(\mathbf{D}) = \sum_{k=1}^{K+1} \sum_{t=D_{(k-1)+1}}^{D_k} |x_t - m_k| \quad (2)$$

変化点時刻ベクトル  $\mathbf{D}$  を求める L1cpd 法の詳細は文献 [3] を参照されたい.

### 3. 実験による評価

本実験では, 2006 年 1 月 1 日から 2016 年 12 月 31 日までの期間における, 平均気圧, 平均気温, 平均湿度, 平均風速, 及び, 日照時間の 5 項目の日別データを用いた<sup>§</sup>. 観測地点としては, 静岡県内の浜松市, 三島市, 及び, 静岡市の 3 地点を選定した. 観測期間の総日数は  $T = 4,018$  となる.

L1cpd 法と L2cpd 法での変化点の違いを定量評価するために次式を用いる [3].

$$H(K) = \frac{1}{2} \sum_{k=1}^K \left( \min_{1 \leq l \leq K} |F_k - D_l| + \min_{1 \leq l \leq K} |D_k - F_l| \right) \quad (3)$$

評価値  $H(K)$  が大きいほど, L1cpd 法と L2cpd 法は互いに異なる時刻を変化点として検出したことを示す. 既存研究 [2] では L2cpd 法において, 最適な変化点数を求めるアルゴリズムが提案されているが, L1cpd 法と L2cpd 法では最適変化点数が異なる. 比較の関係上, 同じ変化点数とする必要があるため, 今回は変化点数  $K = 5$  と設定して実験を行い, 結果を比較した.

表 1 に,  $H(K)$  による定量評価結果を示す. 静岡の湿度に大きい差が見られた. また, 日照はすべての地

表 1: L1cpd 法と L2cpd 法での変化点の違いの定量評価

地点	気圧	気温	湿度	風速	日照
浜松市	999.9	679.6	0.0	131.9	546.7
三島市	0.4	1.0	19.0	216.5	297.8
静岡市	0.0	0.0	1166.0	648.0	437.3

<sup>†</sup>静岡県立大学 経営情報学部

<sup>‡</sup>静岡県工業技術研究所 電子科

<sup>§</sup><http://www.data.jma.go.jp/obd/stats/etrn>

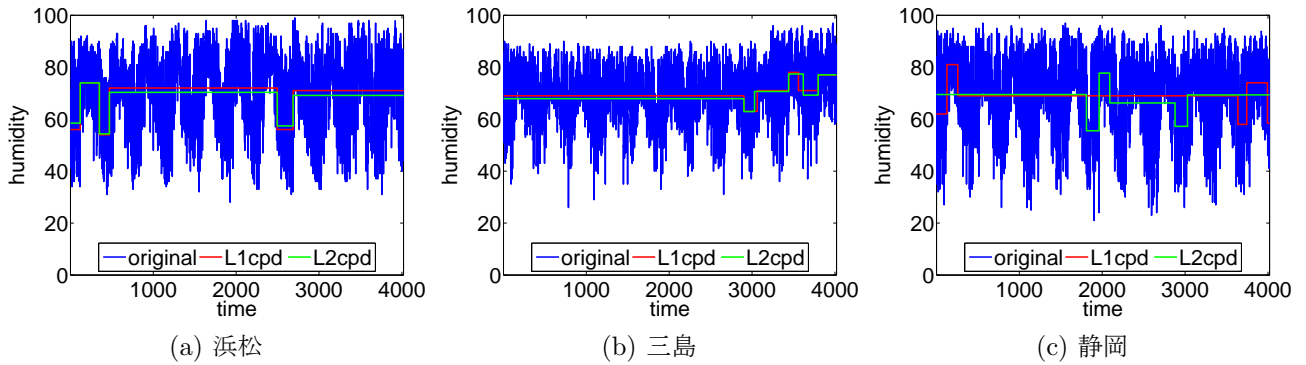


図1: 平均湿度に対し L1cpd 法と L2cpd 法で検出した変化点の比較

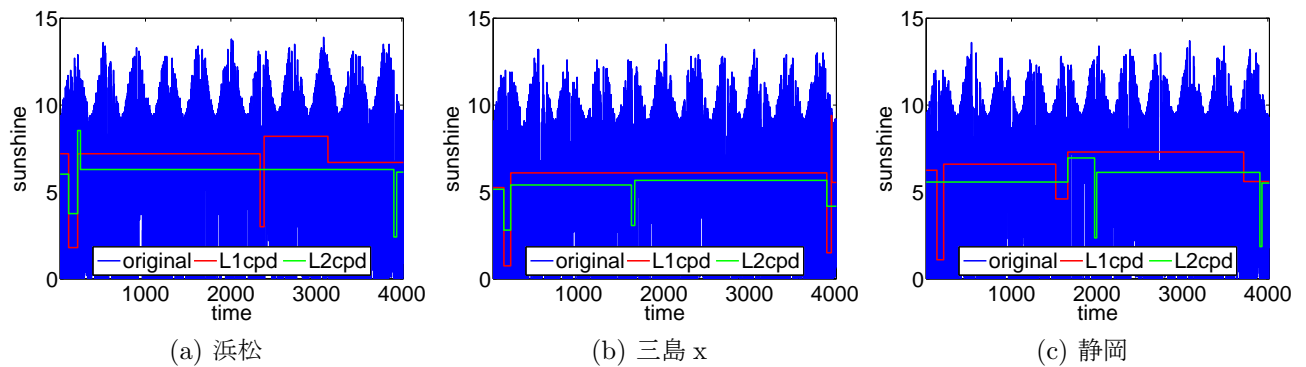


図2: 日照時間に対し L1cpd 法と L2cpd 法で検出した変化点の比較

点においてある程度の差が見られた。そこで、湿度と日照について、詳細な比較を行った。

図1は、 $H(K)$  が最大となる平均湿度に着目し、L1cpd 法と L2cpd 法で検出した変化点を比較したグラフである。各グラフにおいて、横軸は2006年1月1日からの日数、縦軸は平均湿度を表し、青線で測定された湿度、赤線で L1cpd 法で得られた中央値、及び、緑線で L2cpd 法で得られた平均値を示す。図1より、浜松と三島では L1cpd 法と L2cpd 法のどちらでも平均湿度の変化点に大きな違いがみられない結果となった。しかし、静岡の L1cpd 法で検出したデータでは測定した日数の始めの時期と終わりの時期に変化点がみられるのに対し、L2cpd 法で検出したデータでは測定日数の半ばごろに変化点がみられる結果となった。

図2は、 $H(K)$  がどの地点でも大きい日照時間に着目し、L1cpd 法と L2cpd 法で検出した変化点を比較したグラフである。L1cpd 法と L2cpd 法では、多くの検出点が異なる結果となった。L1cpd 法では、2006年春から夏の間、どの場所においても同様の変化点が抽出された。2006年は気象庁の報道発表資料[4]によれば“春から梅雨時期にかけて全国的に日照時間が顕著に少なかった。梅雨明け以降は一転、日照時間は平年を上回ることが多かったが、年間日照時間は全国で少なかった。”とされており、春から梅雨にかけての日照時間減が抽出された可能性がある。

L2cpd 法の静岡においては、2006年春の日照時間の変化点は抽出されなかった。変化点数  $K$  を5としてい

るため、検出されなかった点に変化が無いということでは無いことに注意したい。幾つかの手法を組み合わせ、総合的に評価することが重要であると考えられる。

#### 4. おわりに

本稿では、気温や湿度など農業環境データを用いて、L1cpd 法と L2cpd 法で検出する変化点を比較し、両者の違いを評価した。今後は、さらに多様な環境情報や観測地点での評価実験を進める。

謝辞 本研究は、科学研究費補助金基盤研究(C)(No.15K00429)の助成を受けた。

#### 参考文献

- [1] K. Saito, K. Ohara, M. Kimura and H. Motoda, "Change Point Detection for Burst Analysis from an Observed Information Diffusion Sequence of Tweets," Journal of Intelligent Information Systems (JIIS), Vol.44, Iss.2, pp 243-269, 2015.
- [2] 杉澤 優馬, 伏見 卓恭, 齊藤 和巳, "時系列変化点の異種時系列への影響度分析," 第12回情報科学技術フォーラム (FIT2013), 2013.
- [3] 小林 えり, 伏見 卓恭, 齊藤 和巳, 池田 哲夫, "メディアンに基づく時系列データの変化点検出法," 第13回情報科学技術フォーラム (FIT2014), 2014.
- [4] 気象庁, "2006年(平成18年)の天候," 報道発表資料, 2007年1月4日