

# 差分累積値に基づく農業環境データの可視化分析法

岩崎 清斗<sup>1,a)</sup> 伏見 卓恭<sup>2</sup> 大久保 誠也<sup>3</sup> 齊藤 和巳<sup>3,4,5</sup>

受付日 2018年11月19日, 再受付日 2019年1月7日,  
採録日 2019年1月30日

**概要:** 本研究の目的は, 熟練農業者の活動が環境に与える影響の分析および可視化である. 特に本研究では, 環境情報の時系列データがどのように変化するか注目する. そこで, 差分累積値に基づく農業環境データの可視化分析法を提案する. 提案手法は, 差分の累計値によるモデルに基づいており, 周期的なデータに対しても, 値が全体的に高い傾向にあるのか低い傾向にあるのかを明らかにすることが可能である. また, 中長期的な環境データから, 大きな変化が発生した点を検出できる. 評価実験では, まず静岡県内の4つのバラ農家に, データ収集用小型デバイスを設置し, 各種環境データを収集した. そして, 飽差データに対して提案手法を適用することにより, 変化点の抽出を行った. 実験結果から, 提案手法は環境変化の特徴的な時点を抽出可能なことが明らかとなった.

**キーワード:** 時系列データ分析, 差分累積値法, 変化点検出, 農業環境データ

## Visual Analysis Method of Agricultural Environment Data based on Cumulative Difference Values

KIYOTO IWASAKI<sup>1,a)</sup> TAKAYASU FUSHIMI<sup>2</sup> SEIYA OKUBO<sup>3</sup> KAZUMI SAITO<sup>3,4,5</sup>

Received: November 19, 2018, Revised: January 7, 2019,  
Accepted: January 30, 2019

**Abstract:** As a step to develop a method for automatically acquiring expert knowledge that can contribute to productivity in agriculture, we propose a new visualization method for analyzing long-term changes agricultural environment information. More specifically, in our proposed method, after transforming original periodic environment data into cumulative difference values (CDVs). We visualize them together with some annotations of change points obtained by our change-point detection method. In our experiments, we focused on humidity deficit (HD) as important agricultural environment, which can be computed from some environment variables such as a temperature, and collected them from vinyl greenhouses owned by four rose farmers by setting our original IoT devices. In our experiments, we show that our proposed method can produce some naturally interpretable visualization results.

**Keywords:** periodic environment data, agricultural environment, visualization

<sup>1</sup> 静岡県工業技術研究所機械電子科  
Industrial Research Institute of Shizuoka Prefecture,  
Shizuoka 421-1298, Japan  
<sup>2</sup> 東京工科大学コンピュータサイエンス学部  
School of Computer Science, Tokyo University of Technol-  
ogy, Hachioji, Tokyo 192-0982, Japan  
<sup>3</sup> 静岡県立大学経営情報学部  
School of Management and Information, University of  
Shizuoka, Shizuoka 422-8526, Japan  
<sup>4</sup> 神奈川大学理学部  
Faculty of Science, Kanagawa University, Hiratsuka, Kana-  
gawa 259-1293, Japan  
<sup>5</sup> 理化学研究所革新知能統合研究センター  
Center for Advanced Intelligence Project, RIKEN, Chuo,  
Tokyo 103-0027, Japan

## 1. はじめに

近年, IoT (Internet of Things) 技術が急速に発展している. それにともない, 農業・医療・教育等の専門知識と熟練が不可欠な分野においても, 専門家や熟練者がどのような状況でどのような活動をしているかに関する情報を, 容易に入手可能となってきている. 一方, 特に農業分野では担い手の高齢化による労働力不足が深刻化しており, 作業の合理化や技術の継承が課題となっている.

<sup>a)</sup> [kiyoto1.iwasaki@pref.sizuoka.lg.jp](mailto:kiyoto1.iwasaki@pref.sizuoka.lg.jp)

農業において、温度・湿度・照度等の環境は非常に重要な要素である。熟練農家は、様々な手段を用いて環境を制御することにより、より多くの収穫や高い品質の作物を得ている。しかし、熟練農家の持つ技術が、生産性や環境にどのような影響をどのような形で与えているのかは、明らかになっていない部分が多い。加えて、環境の影響は1日のみでは完結せず、収穫までのある程度の期間に及ぶ。また、環境変化は年・月・日単位で周期的なものである。したがって、農業環境データの分析においては、解析対象がどのような位置づけの日であるのか等を明らかにする必要がある。

本研究の目的は、熟練農家に特有な環境の解析である。特に、農産物の生産性向上に向けて、環境情報オンライン可視化システムのプロトタイプ構築を目指す。本稿では、この目的に向けた基本ステップとして、差分累積値に基づく農業環境データの可視化分析法を提案する。そして、評価実験により、提案手法の有効性を検証する。具体的には、静岡県内にある4件のバラ農家から収集した環境データに提案手法を適用し、その妥当性を評価する。

本稿は以下に示す構成である。2章で提案手法に関連する既存研究について整理する。3章で提案手法の定義と計算アルゴリズムについて解説する。4章で実データを用いた評価実験およびその結果について議論する。最後に本稿のまとめと今後の課題について述べる。

## 2. 関連研究

本研究では、時系列データの新たな可視化分析法を提案するとともに、提案手法を農業環境データに適用する。本章では、時系列データ解析と変化点検出手法、ならびに農業データ解析の関連研究について述べる。

本研究では、農業環境データの変化を検出するため、データ生成の背後に内在する基本メカニズムの変化検出をレジーム切替 (regime switching) 問題 (e.g., 文献 [2], [4]) として定式化し、その変化点検出のために技術 [4] を利用する。この問題設定は、従来から使用されてきた、定常モデルと比較して統計的に有意な短期的外れ値を求める異常検出 (anomaly detection) や、確率分布の混合モデルとして定常モデルを設定する統計的機械学習の枠組み [1] とは、一線を画すものである。従来の異常検出に使用される統計的手法は、与えられたデータに対して統計モデル (インスタンスの大多数は正常であるという仮定) を適合させ、統計的検定によって未知のインスタンスがこのモデルに属するか否かを決定する。このような手法では、適用された統計的検定に基づき、学習モデルから生成される確率が低いインスタンスは異常とされる。一方、本研究では、モデルやメカニズムの変化を検出するため、データ背後に内在する規則性や知識の抽出と親和性が高い。特に、時間で変化するモデルパラメータをレジームスイッチングとして扱うた

め、従来の典型的異常検出技術とは方向性が異なっている。

本研究は、Kleinberg [3] や Swan ら [5] と同様に、時系列データからの構造抽出を回顧的 (retrospective) な枠組みにより実現する。たとえば、Kleinberg の研究では、文書ストリーム内のトピックの出現をバーストとして表現した後、その入れ子構造を推定し、ある期間におけるトピックのアクティビティを要約することで、バースト分析を容易にしている。一方、Swan らの研究は、仮説検定に基づいた時間経過による特徴出現モデルを使用することにより、比較的短期間で出現するコーパス内の主要トピックに対応する情報をクラスターとして生成する。本研究も同様に、過去に起こった現象を理解するという回顧的枠組みを採用する。しかし、レジームスイッチングに基づく変化を仮定し、中長期的な変化の検出も試みるという点で、従来の研究とはモチベーションが異なっている。

各種農業データを取得し、そこから環境制御技術について評価する研究も進められている [11]。たとえば、熟練農家と初心農家の1日分の環境データを時系列順にプロットして比較することや、特定の代表日を定めて環境変化の差について考察すること、環境変化の波形に理由付けをしてモデル化を行うこと等が行われている。これらの研究においても、望ましい環境のモデル化等が行われているが、特定の代表日を扱っており、農作期間全体を扱っていない。一方、山岸ら [6], [7], [8], [9] は、多様な時系列データをレジーム切替問題として分析している。たとえば文献 [8] では、晴れ、曇り、雨等名義変数で記述される時系列データを、多項分布でモデル化し、その顕著なモデル変化点を検出することにより、効果的に可視化している。これに対し、本稿では新たに累積差分値の考え方を採用する。具体的には、植物の生育に大きな影響を及ぼす飽差等の数値変数で記述される時系列データを対象に、その平均値からの差分の累計値をモデル化し、その顕著なモデル変化点を検出することにより、効果的な可視化を行う。

## 3. 提案手法

### 3.1 提案手法概要

提案手法は、周期性を有する時系列データが与えられた際、各時点が全体的な変化のうち、どのような位置づけであるかを明らかにする。また、変化点検出により、どのタイミングで傾向が大きく変わっているかを示す。たとえば、温度データ等は夜は低く日中は高い傾向がある。このようなデータが1カ月分与えられた際、データ中のある期間がどのような変化傾向にあるかや、ある日がどのような変化の中の一部であるか等を明らかにする。

提案手法の特徴は、周期性を有する時系列データの変化傾向を顕著に表現するために、オリジナル時系列データの累積差分値を求めて分析すること、および、累積差分値の時系列データに対し変化点検出法 [4] を適用し、変化傾向

にアノテーションを付与する点にある,

### 3.2 累積差分値によるアノテーション付き可視化法

気温、飽差等の第  $d$  日目の時刻  $h$  における観測値を  $u_{d,h}$  とすれば、このような周期性を有する農業環境時系列データ  $U$  は  $U = \{u_{d,h} \mid d = 1, \dots, D, h = 1, \dots, H\}$  と表せる。ここで、 $D$  は観測日数、 $H$  は 1 日あたりの観測回数である。また、観測間隔はつねに一定であるとする。

前述のとおり、気温等の環境データは、昼間の値は高く、夜間では低い等の周期性を有しており、各時刻において異なる平均値を持つ。そこで、各観測時刻に対し、各時刻における平均からの差分値  $v_{d,h}$  を次式で求める。

$$v_{d,h} = u_{d,h} - \frac{1}{D} \sum_{d'=1}^D u_{d',h}$$

この時系列データ  $\mathcal{V} = \{v_{d,h} \mid d = 1, \dots, D, h = 1, \dots, H\}$  を対象とした分析を、DV (difference value) 分析と呼ぶ。

しかしながら、日々の観測データの変動は比較的大きく、DV 分析では一般に限界がある。そこで、より顕著にデータの変化傾向を表現するために、上述した差分値の累積値  $w_{d,h}$  を次式で求める。

$$w_{d,h} = \sum_{d'=1}^{d-1} \sum_{h'=1}^H v_{d',h'} + \sum_{h'=1}^h v_{d,h'}$$

この時系列データ  $\mathcal{W} = \{w_{d,h} \mid d = 1, \dots, D, h = 1, \dots, H\}$  を対象とした分析を、CDV (cumulative difference value) 分析と呼ぶ。

$w_{d,h}$  に対し時刻順にサフィックスを付与することにより、時系列データ  $\mathcal{X} = \{x_t \mid t = 1, \dots, T\}$  を構成する。明らかに、 $\mathcal{X} = \mathcal{W}$  であり、かつ、 $T = D \times H$  となる。また、平均からの差分の累計値を求めているので、その最終値は  $x_T = w_{D,H} = 0$  となる。CDV 分析では、時系列データ  $\mathcal{X}$  を時刻に対してプロットしたときの、グラフの傾きが重要となる。すなわち、増加傾向ならば平均より高い状態が続き、逆に、減少傾向ならば平均より低い状態が続き、比較的平坦ならば平均的な状態が続いていることが分析できる。

### 3.3 変化点検出法

顕著なデータの変化傾向を表現するため、時系列データ  $\mathcal{X}$  に対して変化点検出法を適用することにより、変化点数  $K$  の近似階段関数  $\mathcal{Z}$  をアノテーションとして求める。そして、時系列データ  $\mathcal{X}$  と階段関数アノテーション  $\mathcal{Z}$  を同時にプロットすることにより、増加傾向・減少傾向・比較的平坦なパートの識別をサポートする。

以下に、処理手順を示す。

入力：オリジナル時系列データ  $U$ 、変化点数  $K$ 。

(1) 時系列データ  $U$  から DV 分析時系列データ  $\mathcal{V}$ 、CDV 分析時系列データ  $\mathcal{W}$  と  $\mathcal{X}$  を求める。

(2) 時系列データ  $\mathcal{X}$  から階段関数アノテーション  $\mathcal{Z}$  を変化点検出法で求める。

出力：時系列データ  $\mathcal{X}$  と階段関数  $\mathcal{Z}$  を同時にプロットした可視化結果。

一般に、累積差分値によるアノテーション付き可視化法においては、時系列データの概要を分析する場合は  $K$  の値は小さく設定し、詳細に分析する場合は大きな値に設定する。以下では、変化点数  $K$  で求めた階段関数アノテーションを CPA (Change-Point Annotation)  $K$  と表記する。

ステップ 2 において、時系列データ  $\mathcal{X}$  から、増加や減少の顕著なパートを変化点として検出する具体的な方法は、以下ようになる。いま、時系列データ  $\mathcal{X}$  に変化点が存在しないとし、1 つの値で近似するならば、その平均値により二乗誤差  $E_0$  は最小化され、次式で求められる。

$$\begin{aligned} E_0 &= \sum_{t=1}^T \left( x_t - \frac{1}{T} \sum_{s=1}^T x_s \right)^2 \\ &= \sum_{t=1}^T x_t^2 - \frac{1}{T} \left( \sum_{s=1}^T x_s \right)^2. \end{aligned}$$

次に、時刻  $A$  が変化点であるとし、その前後をそれぞれ 1 つの値で近似するならば、前後それぞれの区間での平均値により二乗誤差  $E_1$  は最小化され、次式で求められる。

$$\begin{aligned} E_1 &= \sum_{t=1}^A \left( x_t - \frac{1}{A} \sum_{s=1}^A x_s \right)^2 \\ &\quad + \sum_{t=A+1}^T \left( x_t - \frac{1}{T-A} \sum_{s=A+1}^T x_s \right)^2 \\ &= \sum_{t=1}^T x_t^2 - \frac{1}{A} \left( \sum_{s=1}^A x_s \right)^2 - \frac{1}{T-A} \left( \sum_{s=A+1}^T x_s \right)^2 \end{aligned}$$

いま、時系列データ  $\mathcal{X} = \{x_1, \dots, x_T\}$  を次式により  $\mathcal{Y} = \{y_1, \dots, y_T\}$  へと変形する。

$$y_t = \sum_{s=1}^t x_s \quad t \in \{1, \dots, T\}.$$

このとき、二乗誤差  $E_1$  は次式で求められる。

$$E_1 = \sum_{t=1}^T x_t^2 - \frac{1}{A} y_A^2 - \frac{1}{T-A} (y_T - y_A)^2$$

以下では、このような誤差関数を一般化する。変化点の個数は  $K$  個であり、それぞれの変化点は古い順に  $F(1)$  から  $F(K)$  であるとする。また便宜上  $F(0) = 0$  かつ  $F(K+1) = T$  と設定し、 $(K+2)$  個の要素からなる集合  $\mathcal{F} = \{F(0), \dots, F(K+1)\}$  を考える。すると、二乗誤差  $E_K(\mathcal{F})$  は次式で求められる。

$$E_K(\mathcal{F}) = \sum_{t=1}^T x_t^2 - \sum_{k=1}^{K+1} \frac{1}{\Delta F(k)} (y_{F(k)} - y_{F(k-1)})^2.$$

ここで、 $\Delta F(k) = F(k) - F(k-1)$  である。我々の変化点

検出問題は、 $E_K(\mathcal{F})$  を最小化する  $\mathcal{F}$  を求める問題として定式化される。その探索には、文献 [4] と同様に、逐次的に変化点を求める方法を採用する。具体的には、階段関数を定義する  $z_t$  を次式で求める。

$$z_t = \frac{1}{\Delta F(k)}(y_{F(k)} - y_{F(k-1)}).$$

これより、階段関数アノテーションを  $\mathcal{Z} = \{z_t \mid t \in \mathcal{F}\}$  として得ることができる。

実際に提案手法を農業環境データに適用するには、 $K$  の値を目的に合わせて設定する必要がある。たとえば、次のような方法が考えられる。目的の 1 つは各ハウスの比較であるので、最初は  $K$  の値を 2 のような小さな値に設定することで、大きな変化を起こしている点が異なっているか否かを検討する。もし、特徴的なハウスが見つかった場合、より詳細な分析を行うために  $K$  の値をより大きい 8 のような値に設定して解析を行う。このように、 $K$  の値を徐々に大きくしていくことにより、各ハウスの特徴を明らかにしていく。

## 4. 評価実験

### 4.1 評価実験概要

評価実験により、提案手法の妥当性を評価する。具体的には、1) 農業環境データを収集するための小型デバイスを開発し、2) 静岡県内にある 4 件のバラ農家のビニールハウスに設置することで、環境データの収集を行い、3) 得られたデータに対して提案手法を適用し解析を行うことにより、提案手法の有効性を評価する。特に本稿では、各種データのうち、特に飽差を対象とした解析を行う。飽差とは、ある温度と湿度の空気に、あとどれだけ水蒸気の入る余地があるかを示す指標であり、農産物の育成に、ある一定範囲での飽差制御が重要であるとされている [10]。

### 4.2 データ収集法

環境データを収集するための小型デバイスを開発した。開発の際、いくつかの点に配慮した。たとえば、農家の活動によってはデバイスを一時撤去する必要がある可能性があること、研究終了後にはデバイスを撤去する必要があること、設置するデバイスが環境に大きな変化を与えると従来からある知見が得られないこと等がある。これらのことをふまえ、設置・再設置が容易であり、かつ既存の設備に影響を与えない小型デバイスとして設計を行った。開発した小型デバイスは、RaspberryPi3 と各種センサからなる。小型デバイスの詳細を表 1 に示す。本デバイスは、おおよそ 5 分に 1 回の頻度で各種情報の収集を行う。

小型デバイスを設置したのは、静岡市内にある 4 つの農家における、バラのビニールハウスである。各農家のビニールハウスは、詳細は異なるが、環境を制御する手段を備えている。たとえば、温度等を調整する暖房や、日射量

表 1 環境データ収集用デバイス詳細

Table 1 Details of the device to collect environmental data.

センサ名	型番
本体	RaspberryPi3 ModelB
温度センサ	BME280
湿度センサ	BME280
照度センサ	TSL2561
カメラ	KP-CAM32




表 2 観測したデータと補間値

Table 2 Observed data and interpolation value.

日付	湿度	温度	観測/補間
18/03/29 16:00:00	31.92	31.83	補間値
18/03/29 16:01:42	32.34	31.61	観測値
18/03/29 16:05:00	32.18	31.61	補間値
18/03/29 16:06:42	32.10	31.61	観測値
18/03/29 16:10:00	32.36	31.39	補間値
18/03/29 16:11:42	32.50	31.28	観測値
18/03/29 16:15:00	32.79	31.02	補間値
18/03/29 16:16:42	32.94	30.88	観測値
18/03/29 16:20:00	33.28	30.58	補間値
18/03/29 16:21:42	33.45	30.43	観測値
18/03/29 16:25:00	33.62	30.56	補間値
18/03/29 16:26:42	33.70	30.62	観測値

等を調整するカーテン等がある。ビニールハウス内の環境は、これらの設備が自動もしくは手動で動作することにより、制御されている。ただし、実験期間中、農家 A (ハウス A) のみは、出荷期間でなかったため、これらの設備はほとんど使用されていない。

4 つの農家に共通したデータ収集期間は 2018 年 3 月 26 日 16 時 10 分から 2018 年 5 月 8 日 9 時 55 分までであり、本解析では 2018 年 3 月 27 日 0 時 0 分から 2018 年 5 月 7 日 24 時 00 分までのデータを使用した。ここで、各データは、おおよそ 5 分間隔であり、厳密に 5 分間隔ではない。また、電源投入時からデータが収集されるため、各農家のデータ収集時刻に差が存在する。そこで、2018 年 3 月 27 日 0 時 0 分から 5 分間隔のデータを、補間することによって作成した。補間は Python の Pandas を用いた。補間方法は index であり、時間に応じた按分を行っている。最終的なデータは、1 日あたり 288 次元ベクトルで、それが 42 日分である。すなわち、 $D = 42$  かつ  $H = 288$  である。観測したデータと補間したデータの例を表 2 に示す。観測値と書かれているデータが、実際に小型デバイスで観測した値である。また、補間値と書かれたデータが追記した値である。ハウス間の比較を行うには、同じ時刻に観測を行うことが望ましいが、設計上ずれてしまうため、時刻による線形補間を行うことで、5 分置ききのデータを追記している。

また、小型デバイスは飽差 HD を直接測定するセンサを備えていないため、以下の式により求めた。

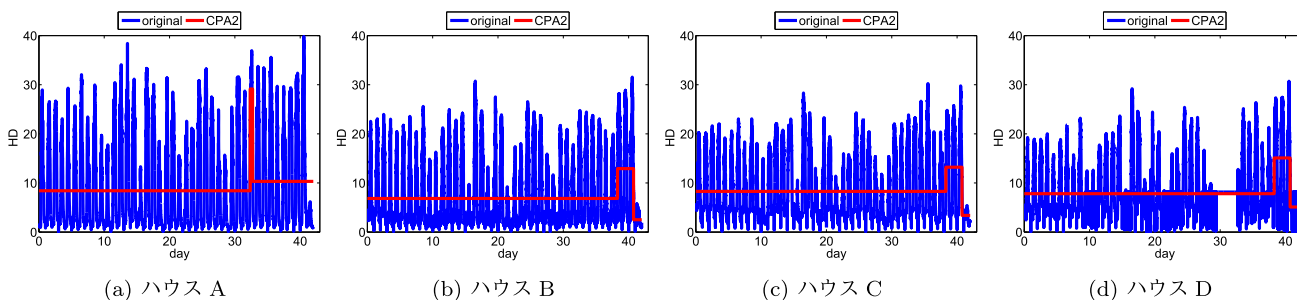


図 1 飽差 (HD) のオリジナル時系列に対する CPA2 の適用

Fig. 1 Result of applying CPA2 to the original time series data of Humidity Deficit.

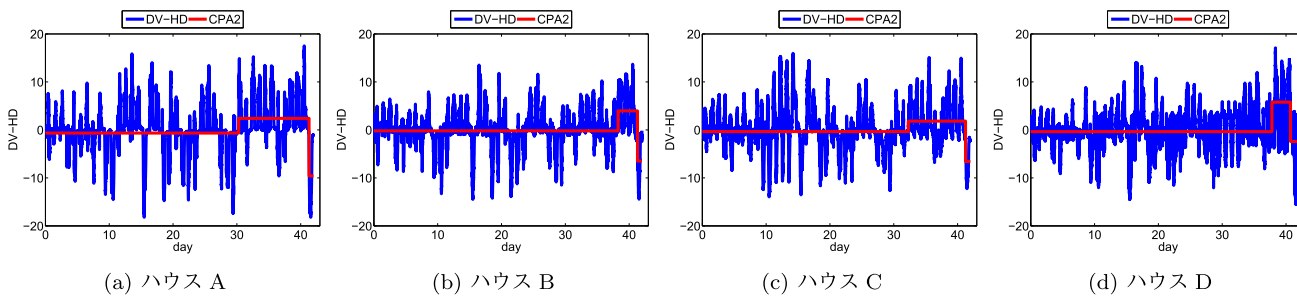


図 2 飽差 (HD) の平均差分値 (DV) 時系列に対する CPA2 の適用

Fig. 2 Result of applying CPA2 to the difference value time series data of Humidity Deficit.

$$HD = (100 - H) * \frac{217 * \frac{6.1078 * 10^{7.5 * T}}{T + 237.3}}{T + 273.15}$$

ここで、 $H$  は湿度、 $T$  は気温である．一般には、飽差の値は  $3 \sim 6 \text{ g/m}^3$  がよいとされている [10].

### 4.3 可視化結果

オリジナル時系列データ  $U$ 、平均差分値時系列データ  $V$ 、累積差分値時系列データ  $W$  について、可視化を行った。

オリジナル時系列データならびに CPA2 による変化点検出結果を、図 1 に示す。青い線が時系列データ、赤い線がデータの変化点を表している。時系列データは飽差の値そのものである。すべての農家において、測定期間終盤に変化点が検出されている。一方で、時系列データは激しく変動しており、一見しただけでは、変化点等の特徴を掴むことは難しい。つまり、CPA2 による変化点は検出されているものの、どのような変化が起きているのかを直感的に理解できる可視化結果となっていない。

平均差分値 (DV) 時系列データならびに CPA2 による変化点検出結果を、図 2 に示す。時系列データは平均差分値であるため、平均値よりも高いか低いかを表している。したがって、0 よりも高ければ飽差は高い値であることを意味し、0 よりも低ければ飽差は低い値であることを意味する。このデータにおいても、すべての農家において、測定期間終盤に変化点が検出されている。また、ハウス A においては、30 日目から 41 日目の期間に平均より高めの値が続いており、そのために“開始から 29 日目”、“31 日目から 41 日目”、“42 日目以降”の 3 つに分類する変化点検

出結果となったことが分かる。

平均差分値時系列データの可視化結果は、どの期間に高い値をとっていたかや、どのような理由で変化点とされたかを把握しやすい。一方、時系列データの値は正負の値を振動しており、全体的に高い傾向にあるのか低い傾向にあるのかを把握するのは難しい。たとえば、ハウス A の 0～10 日目までの期間は振動しており、増加傾向にあるのか減少傾向にあるのかを把握できる可視化結果となっていない。

累積差分値 (CDV) 時系列データならびに CPA2 による変化点検出結果を、図 3 に示す。時系列データは累積差分値であるため、各値は 0 日目からの累積差分を表している。そして、時系列データのグラフの傾きが、平均よりも高い値なのか低い値なのかを表している。つまり、傾きが正に大きければ、平均値よりも高い値が連続している期間であり、傾きが負に大きければ、平均値よりも低い値が連続している期間である。また、変化点は、大きな傾きとなる期間の中心部分を表す。このデータにおいては、どのハウスにおいても、20 日目前後と 35 日目前後に変化点が検出された。実際に、飽差のグラフの傾きは 20 日目前後では負に、35 日目前後では正に大きい。また、すべてのハウスでグラフの形状は類似しているが、ハウス B と C の変化量は比較的小さく、ハウス A と D の変化量は比較的大きい。このことから、ハウス B と C は、比較的安定した環境に制御されていることが分かる。

累積差分値時系列データの可視化結果は、飽差の具体的な値は把握しにくい。一方、長期的に見て飽差が高い値をとる傾向があるのか、低い値をとる傾向があるのかを把握

しやすい。たとえば、ハウス A の 0~10 日目までの期間は、飽差の値は振動しているが、全体としては平均よりも低い値をとっていることが分かる。変化点の検出においては、検出結果のみからは何らかの期間を示すことはできないが、変化点の前後の時系列データをあわせて見ることで、重要な変化をしている期間を把握することができる。

累積差分値 (CDV) 時系列データならびに CPA8 による変化点検出結果を、図 4 に示す。変化点を 8 つ検出するた

め、平均値より値が大きい期間と小さい期間の中心となる日を検出することができている。

### 5. 考察

図 1 と図 2 では、比較的似た時点が変化点として検出されているが、図 3 では異なる時点が検出されている。これは、前者では値そのものが評価されているのに対し、後者では値の変化が評価されているためであると考えられる。

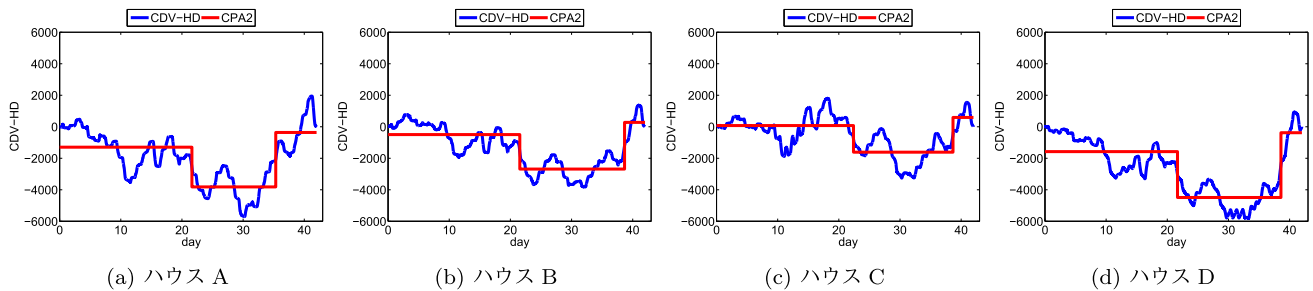


図 3 飽差 (HD) の累積差分値 (CDV) 時系列に対する CPA2 の適用

Fig. 3 Result of applying CPA2 to the cumulative difference value time series data of Humidity Deficit.

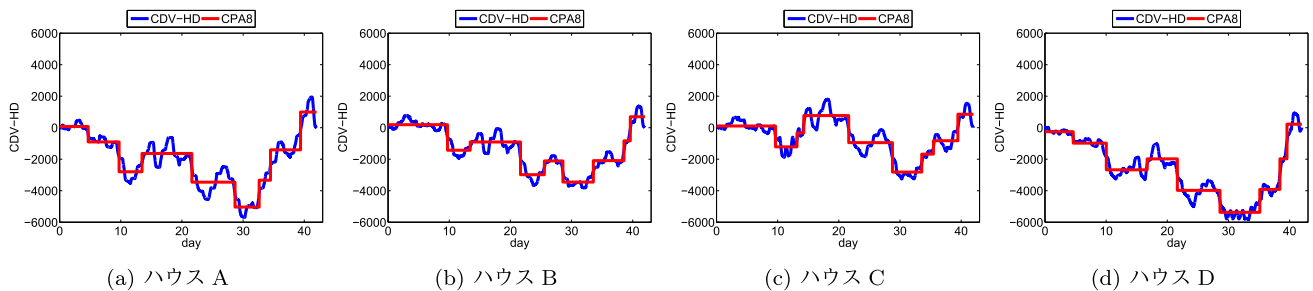


図 4 飽差 (HD) の累積差分値 (CDV) 時系列に対する CPA8 の適用

Fig. 4 Result of applying CPA8 to the cumulative difference value time series data of Humidity Deficit.

表 3 各ハウスからのコメント

Table 3 Comments from each house.

番号	ハウス	コメント
1	ハウス A	長期間の変化の傾向が分かることで、年間通じて安定的な生産計画を立てることができる。
2	ハウス A	季節の変わり目で環境制御方法を経験的に変更させているが、気象条件等で変わり目が前後する場合があるため、その変わり目の精度を上げるために中長期的な傾向が分かると良い。
3	ハウス B	年間通じての飽差傾向が分かれば、来年以降の時季ごとの対策が立てられる。
4	ハウス B	平均からの差分ではなく、飽差の最適値 (3~6 g/m <sup>2</sup> ) からの差分が分かると現在値からどれだけ修正すればよいか分かる。
5	ハウス B	差分累積値のグラフから変化度合いの大きさが読み取れるが、実際の数値でどれくらい変化しているのか把握したい。
6	ハウス C	飽差がどのように植物に作用しているのか正確に分かっていない。飽差を見てどのように制御すればいいか分からない。一方、なるべく安定的な環境にできれば良いという観点から、グラフ結果は活用できる気がする。
7	ハウス C	変化が起きたときにその原因 (温度、湿度どちらが変化したか) が分かると対策が立てやすい。
8	ハウス C	外気との関係性が知りたい。3~5 月は天窓しか制御していないため、外気の影響を受けにくい。どれだけ環境制御ができていたかを比較できる。
9	ハウス D	元々経験則でやっている農家にとっては、こういったグラフが見やすいかということが分からない。逆をいえば、直感的に理解できるグラフであれば見やすいと感じることができる。
10	ハウス D	ハウスごとに傾向の比較ができるのはありがたい。
11	ハウス D	各ハウスの設備でどれだけ完璧に環境制御を行うことができているかは分からない。分かるようになりたい。
12	ハウス D	前の季節に引込まれる影響 (例: 長雨が続き日射量が不十分、夏季が高温過ぎる) を推測できるのは良い。
13	ハウス D	基本値、代表値からどれくらいの差分があるのかを直感的に把握できるようにしたい。

また、図1と図2では、飽差の値が大きい期間や小さい期間が直接的に検出されている。それに対し、図3では、大きく飽差の値が変化する代表点が検出されており、何かの期間は検出されていない。一方で、図1と図2から変化の度合いを読み取ることは難しいが、図3からは読み取ることができる。図3で大きく変化する時点を検出し、図2で変化を起こしている期間を推定する等、2つの分析を組み合わせることで、モデル化に必要な各種データを得ることができると考えられる。

図1と図2の時系列データのグラフから、農家間の類似性を把握することは難しい。一方、図3では、“ハウスAとD”・“ハウスBとC”の2つに分けることができる。比較的増減の小さい“ハウスBとC”は、飽差管理の点において何らかの環境制御が作用していることが考えられる。しかし、環境制御を行っていないハウスAは増減が大きく、それに類似するハウスDは飽差管理の点において改善の余地があると考えられる。

以上のことから、提案手法は従来にはない形によるデータ分析ができており、熟練農家における環境データのモデル化に有用であると考えられる。

## 6. 定性的評価

評価実験により、提案手法が従来にはない形によるデータ分析が可能であることは明らかになった。しかしながら、提案手法が実際の農家で有用であるか否かは明らかではない。そこで、各ハウスの方々から、提案手法による可視化結果に対してコメントをいただいた。各ハウスからのコメントを表3に示す。

全体を通じて、長中期にわたるデータを可視化することについては、肯定的な意見が得られた(コメント1, 3, 12)。時節ごとの特徴をとらえることで生産計画を立てたり、前の時節の影響をとらえることができるようになるため、提案手法による可視化は有用であるといえる。

一方で、本データのみではとらえることができない各要素について、多くの要望が出された。特に、提案手法は傾向を示すのみであり、実際の値は可視化していないため、変化の度合いや各時点での値を得たいという要望が多かった(コメント4, 5, 13)。また、何が原因で変化が起きているのか、その原因、操作の影響を知りたいという要望も多かった(コメント2, 7, 8, 11)。

以上のことから、提案手法による可視化は、実際に農家の役に立つデータを提供できていることが明らかとなった。一方で、提案手法のみでは不十分であり、より多くのデータを可視化する手法が必要であることも明らかとなった。

## 7. おわりに

本研究では、農業環境データを分析する手法を提案した。提案手法は、データの累積差分をとることにより、データ

の変化を特徴付け、中長期的な変化の特徴を明らかにする。提案手法の有効性を示すため、小型デバイスにより静岡市内のビニールハウスのデータを収集し、得られたデータに提案手法を適用した。結果、提案手法の特徴や、各農家が環境を制御した影響等が明らかとなった。今後の課題として、小型デバイスの改良や、熟練農家の技術が環境に与える影響の評価等があげられる。また、本実験では、農家の方が実際に役に立つと感じる、適切な $K$ の決め方は明らかとなっていない。 $K$ の値の適切な設定の仕方も今後の課題である。

謝辞 本研究は、JSPS 科研費 (C) (No.18K11441) の助成を受けたものである。

## 参考文献

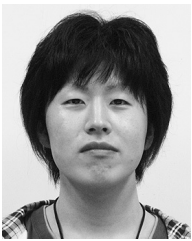
- [1] Chandola, V., Banerjee, A. and Kumar, V.: Anomaly Detection: A Survey, *ACM Comput. Surv.*, Vol.41, No.3, pp.15:1-15:58 (online), DOI: 10.1145/1541880.1541882 (2009).
- [2] Kim, C.-J., Piger, J. and Startz, R.: Estimation of Markov regime-switching regression models with endogenous switching, *Journal of Econometrics*, Vol.143, No.2, pp.263-273 (2008).
- [3] Kleinberg, J.: Bursty and Hierarchical Structure in Streams, *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '02*, pp.91-101, ACM (online), DOI: 10.1145/775047.775061 (2002).
- [4] Saito, K., Ohara, K., Kimura, M. and Motoda, H.: Change point detection for burst analysis from an observed information diffusion sequence of tweets, *J. Intell. Inf. Syst.*, Vol.44, No.2, pp.243-269 (online), DOI: 10.1007/s10844-013-0283-2 (2015).
- [5] Swan, R. and Allan, J.: Automatic Generation of Overview Timelines, *Proc. 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '00*, pp.49-56, ACM (online), DOI: 10.1145/345508.345546 (2000).
- [6] Yamagishi, Y., Okubo, S., Saito, K., Ohara, K., Kimura, M. and Motoda, H.: A Method to Divide Stream Data of Scores over Review Sites, *PRICAI 2014: Trends in Artificial Intelligence - Proc. 13th Pacific Rim International Conference on Artificial Intelligence*, pp.913-919 (online), DOI: 10.1007/978-3-319-13560-1.78 (2014).
- [7] Yamagishi, Y. and Saito, K.: Visualizing Switching Regimes Based on Multinomial Distribution in Buzz Marketing Sites, *Foundations of Intelligent Systems - Proc. 23rd International Symposium, ISMIS 2017*, pp.385-395 (online), DOI: 10.1007/978-3-319-60438-1.38 (2017).
- [8] 山岸祐己, 岩崎清斗, 斉藤和巳: 多群出現順位統計量に基づく時系列データの変換, 情報処理学会論文誌数理モデル化と応用 (TOM), Vol.11, No.1, pp.45-52 (2018).
- [9] 山岸祐己, 斉藤和巳: レビューサイトにおける多項分布に基づくレジームスイッチング検出手法と可視化への応用, 情報処理学会論文誌データベース (TOD), Vol.10, No.3, pp.58-67 (2017).
- [10] 農文協 (編): 野菜 vol.7: もっと知りたい環境制御技術— 日中 CO<sub>2</sub> 濃度, 飽差, 葉面積を管理する, 農山漁村文化協会 (2014).
- [11] 農文協 (編): 野菜 vol.8 特集: ここまで見えた環境制御

技術, 農山漁村文化協会 (2015).



岩崎 清斗

静岡県工業技術研究所機械電子科主任研究員。2011年法政大学工学部システム制御工学科卒業。センサネットワークの研究に従事。



伏見 卓恭 (正会員)

2011年静岡県立大学大学院経営情報学研究科修士課程修了。2014年同大学院経営情報イノベーション研究科博士後期課程修了。同年静岡県立大学大学院経営情報学部客員研究員。2015年筑波大学図書館情報メディア系特別研究員 (PD)。2017年より東京工科大学コンピュータサイエンス学部助教。複雑ネットワーク, 可視化の研究に従事。博士 (学術)。人工知能学会, 日本データベース学会各会員。



大久保 誠也 (正会員)

2000年電気通信大学電気通信学部卒業。2002年同大学大学院電気通信学研究科博士前期課程修了。2005年同大学院電気通信学研究科博士後期課程修了。同年電気通信大学研究員 (COE)。2008年静岡県立大学経営情報学部助教。2017年静岡県立大学経営情報学部講師。現在に至る。量子計算ならびにゲーム情報学の研究に従事。電子情報通信学会, 日本データベース学会各会員。



斉藤 和巳 (正会員)

1963年生。神奈川大学理学部教授。1985年慶應義塾大学理工学部卒業。同年NTT入社, 2007年静岡県立大学経営情報学部教授。1998年東京大学博士 (工学)。複雑ネットワークの研究に従事。電子情報通信学会, 人工知能学会, 日本神経回路学会, 日本応用数理学会, 日本行動計量学会, 日本データベース学会各会員。