

# 多項分布型レジームスイッチング検出手法 による環境情報の可視化

山岸 祐己<sup>1,a)</sup> 岩崎 清斗<sup>2</sup> 齊藤 和巳<sup>1,b)</sup>

**概要:** 状態変化をともなう環境情報は、変化が複雑な場合、無駄な情報量が多く、そのまま可視化しても解釈が難しいため、環境情報の圧縮や単純化を目的とした技術は重要であると言える。よって、本論文では、環境情報はなんらかの理由で時間と共に変化していると考え、その傾向変化をレジームスイッチングに基づくタイムラインとして表現する手法を提案する。提案手法は、各レジームにおける環境情報は多項分布に従っていると仮定し、観測された環境情報における尤度を最大化することによって、それらのモデルパラメータとスイッチング時刻を推定する。提案手法が生成するタイムラインは、非常に単純であるが、情報の本質は失っていないことを現実データによる実験で示す。

**キーワード:** 時系列データ, ストリームデータ, レジームスイッチング, 最尤推定

## Visualization of Environmental Information by Regime Switching Detection Method Based on Multinomial Distribution

YUKI YAMAGISHI<sup>1,a)</sup> KIYOTO IWASAKI<sup>2</sup> KAZUMI SAITO<sup>1,b)</sup>

### 1. はじめに

状態変化をともなう環境情報は、変化が複雑な場合、無駄な情報量が多く、そのまま可視化しても解釈が難しいため、それら環境情報の圧縮や単純化を目的とした技術は重要であると言える。今回扱うような時系列データ（ストリームデータ）の研究では、現時点の状況解析や将来予測に焦点を当てているものもあるが、今回の研究内容は、過去に何が起き、どのような変化をしていたかということに焦点を当てた研究 [1], [2] と類似する。本研究では、レジームスイッチングの検出問題を定式化し、推定されたレジームスイッチングに基づいた時系列データのタイムライ

ンを生成する手法を提案する。ここで、各レジームにおける環境情報の基本生成パターン多項分布に従っていると仮定し、スイッチングが起こるタイムステップとモデルパラメータは、観測された時系列データの尤度を最大化することによって推定する。

本研究は、Kleinberg [1] や Swan と Allan [2] と同様に、回顧的 (retrospective) な枠組みによる時系列データからの構造抽出を目的としている。たとえば、Kleinberg の研究は、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、その入れ子構造を推定することによって、ある期間におけるトピックのアクティビティを要約し、それらの分析を容易にしている。この Kleinberg の手法は、バーストが自然に状態遷移として現れる隠れマルコフモデルを使用しており、電子メールメッセージの階層構造を識別することができる。環境情報における適応を考えると、観測時刻の間隔（データの時間密度）が変化しているものについては、既存のバースト検出技術 [1] とともに、ウィンドウに基づく手法 [3] や複数ストリームを対象とした手

<sup>1</sup> 静岡県立大学  
University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

<sup>2</sup> 静岡県工業技術研究所  
Industrial Research Institute of Shizuoka Prefecture, 2078 Makigaya, Aoi-ku, Shizuoka 421-1221, Japan

a) yamagissy@gmail.com

b) k-saito@u-shizuoka-ken.ac.jp

法 [4] なども適応可能であるが、観測時刻の間隔がほぼ一定のものについては、これら既存手法の有効性は低いことが予想される。さらに、既存のバースト検出技術は、単一情報のバーストを検出するものであり、複数情報とその分布の変化に着目していないため、状態変化などの複数情報の傾向変化を検出することには向いていない。一方、Swan と Allan 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し、コーパス内の主要トピックに対応する情報をクラスタとして生成することに成功している。本研究も同様に、過去に起こった現象を理解するという目的を持っているが、あくまでレジームスイッチングに基づく変化を仮定しているため、このような研究のモチベーションとも離れている。

ここで、今回扱うようなレジームスイッチング検出は、ノベルティ検出や外れ値検出 [5] で使用される技術のような、機械学習の分野で広く研究されている異常検出や変化点検出の典型的技術とは大きく異なることを強調しておく。たとえば、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ものである。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。本研究は、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っているため、これら典型的異常検出技術とは方向性が異なる。同様の方向性を持つ従来アプローチとしては、経済分野におけるレジームスイッチングモデルの研究 [6] があげられるが、これらの研究はガウシアンモデルに大きく依存している。意思決定支援の分野でも、オンラインレビューシステムにおける不正な評価を検出するための技術 [7] がいくつか開発されているが、これらの方法は明確に異常検出技術の領域に分類される。

## 2. 提案手法

### 2.1 問題設定

環境情報の時系列データを  $\mathcal{D} = \{(s_1, t_1), \dots, (s_N, t_N)\}$  とする。ここで、 $s_n$  と  $t_n$  は、 $J$  カテゴリの環境状態と  $n$  番目の観測時刻をそれぞれ表す。 $|\mathcal{D}| = N$  を観測数とすると、 $t_1 \leq \dots \leq t_n \leq \dots \leq t_N$  となる。 $n$  はタイムステップとし、 $\mathcal{N} = \{1, 2, \dots, N\}$  をタイムステップ集合とする。また、 $k$  番目のレジームの開始時刻を  $T_k \in \mathcal{N}$ 、 $\mathcal{T}_K = \{T_0, \dots, T_k, \dots, T_{K+1}\}$  をスイッチングタイムステップ集合とし、便宜上  $T_0 = 1$ 、 $T_{K+1} = N + 1$  とする。すなわち、 $T_1, \dots, T_K$  は推定される個々のスイッチングタイムステップであり、 $T_k < T_{k+1}$  を満たすとする。そして、 $\mathcal{N}_k$  を  $k$  番目のレジーム内のタイムステップ集合とし、各  $k \in \{0, \dots, K\}$  に対して  $\mathcal{N}_k = \{n \in \mathcal{N}; T_k \leq n < T_{k+1}\}$

のように定義する。なお、 $\mathcal{N} = \mathcal{N}_0 \cup \dots \cup \mathcal{N}_K$  である。

いま、各レジームの状態分布が  $J$  カテゴリの多項分布に従うと仮定する、 $\mathbf{p}_k$  を  $k$  番目のレジームにおける多項分布の確率ベクトルとし、 $\mathcal{P}_K$  はそれら確率ベクトルの集合、つまり  $\mathcal{P}_K = \{\mathbf{p}_0, \dots, \mathbf{p}_K\}$  とすると、 $\mathcal{T}_K$  が与えられたときの対数尤度関数は以下のように定義できる。

$$L(\mathcal{D}; \mathcal{P}_K, \mathcal{T}_K) = \sum_{k=0}^K \sum_{n \in \mathcal{N}_k} \sum_{j=1}^J s_{n,j} \log p_{k,j}. \quad (1)$$

ここで、 $s_{n,j}$  は  $s_n \in \{1, \dots, J\}$  を

$$s_{n,j} = \begin{cases} 1 & \text{if } s_n = j; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

のように変換したダミー変数である。各レジーム  $k = 0, \dots, K$  と各状態  $j = 1, \dots, J$  に対する式 (1) の最尤推定量は  $\hat{p}_{k,j} = \sum_{n \in \mathcal{N}_k} s_{n,j} / |\mathcal{N}_k|$  のように与えられる。これらの推定量を式 (1) に代入すると以下の式が導ける。

$$L(\mathcal{D}; \hat{\mathcal{P}}_K, \mathcal{T}_K) = \sum_{k=0}^K \sum_{n \in \mathcal{N}_k} \sum_{j=1}^J s_{n,j} \log \hat{p}_{k,j}. \quad (3)$$

したがって、スイッチングタイムステップの検出問題は、式 (3) を最大化する  $\mathcal{T}_K$  の探索問題に帰着できる。

しかし、式 (3) だけでは  $\mathcal{T}_K$  の導入によってどれだけ尤度が改善したかという直接的な評価をすることができない。この問題において、レジームスイッチングを考慮しないときの尤度からの改善度合いを評価することは重要であるため、尤度比最大化問題として目的関数を構築し直す。もし、レジームスイッチングのような変化が存在しない、すなわち  $\mathcal{T}_0 = \emptyset$  と仮定すると、式 (3) は

$$L(\mathcal{D}; \hat{\mathcal{P}}_0, \mathcal{T}_0) = \sum_{n \in \mathcal{N}} \sum_{j=1}^J s_{n,j} \log \hat{p}_{0,j}, \quad (4)$$

となる。ここで、 $\hat{p}_{0,j} = \sum_{n \in \mathcal{N}} s_{n,j} / N$  である。よって、 $K$  個のスイッチングを持つ場合と、スイッチングを持たない場合の対数尤度比は

$$LR(\mathcal{T}_K) = L(\mathcal{D}; \hat{\mathcal{P}}_K, \mathcal{T}_K) - L(\mathcal{D}; \hat{\mathcal{P}}_0, \mathcal{T}_0). \quad (5)$$

のように与えられる。最終的に、この問題は上記の  $LR(\mathcal{T}_K)$  を最大化する  $\mathcal{T}_K$  の探索問題に帰着できる。

### 2.2 解法

式 (5) を網羅的に解くと最適解が保証されるが、計算量が  $O(N^K)$  となってしまうため、ある程度大きい  $N$  に対して  $K \geq 3$  となってしまうと、実用的な計算時間で解くことができない。したがって、我々は任意の  $K$  について解くための高速な解法を提案する。以下では、まず貪欲法 (A1) と局所探索法 (A2) を説明し、更にそれらを組み合わせた提案解法について説明する。

### 2.2.1 貪欲法

まず、貪欲法 (A1) の手順について説明する。このアルゴリズムは、バックトラッキングをしないデータの2分割の繰り返しである。つまり、既に選択された  $(k-1)$  個のスイッチングタイムステップ  $\mathcal{T}_{k-1}$  を固定したまま  $k$  番目のスイッチングタイムステップ  $T_k$  を  $\mathcal{T}_{k-1}$  に新たに追加することを繰り返す。また、アルゴリズムの終了条件として最小記述長原理 (MDL) を採用する。貪欲法アルゴリズムの手順は以下となる。

- A1-1.**  $k = 1, \mathcal{T}_0 = \emptyset$  のように初期化する。
- A1-2.**  $T_k = \arg \max_{t_n \in \mathcal{T}} \{LR(\mathcal{T}_{k-1} \cup \{t_n\})\}$  を探索する。
- A1-3.**  $\mathcal{T}_k = \mathcal{T}_{k-1} \cup \{T_k\}$  のように更新する。
- A1-4.** もし  $-L(\mathcal{D}; \hat{\mathcal{P}}_k, \mathcal{T}_k) + (J-1)k \log N/2 > -L(\mathcal{D}; \hat{\mathcal{P}}_{k-1}, \mathcal{T}_{k-1}) + (J-1)(k-1) \log N/2$  なら  $\mathcal{T}_{k-1}$  を  $\mathcal{T}_K$  として出力して終了する。
- A1-5.**  $k = k+1$  とし、A1-2 に戻る。

ここで、A1-3 での  $\mathcal{T}_k$  の各スイッチングタイムステップは、 $T_{k-1} < T_k$  を満たすように再インデックスする。明らかに、このアルゴリズムの計算量は  $O(NK)$  と高速であるため、大規模な  $N$  に対しても実用的な計算時間で結果を得ることが可能である。しかし、先ほども説明したように、このアルゴリズムはバックトラッキングを行わないため、プアーな局所解に陥ってしまうことが危惧される。

### 2.2.2 局所探索法

次に、局所探索法 (A2) について説明する。このアルゴリズムは、A1 で得られた解  $\mathcal{T}_K$  から始まり、スイッチングタイムステップの改善を1つずつ試みるものである。つまり、 $k$  番目のスイッチングタイムステップ  $T_k$  を一度取り去り、残った  $\mathcal{T}_K \setminus \{T_k\}$  を固定して、よりよい尤度を得られる  $T'_k$  を探索することを  $k = 1$  から  $K$  まで繰り返す。ここで、 $\cdot \setminus \cdot$  は集合差を表す。もし、すべての  $k$  ( $k = 1, \dots, K$ ) に対してスイッチングタイムステップの置換が行われない、すなわち、すべての  $k$  に対して  $T'_k = T_k$  ならば、これ以上の改善は望めないとして処理を終了する。局所探索法のアルゴリズムは以下となる。

- A2-1.**  $k = 1, h = 0$  のように初期化する。
- A2-2.**  $T'_k = \arg \max_{t_n \in \mathcal{T}} \{LR(\mathcal{T}_K \setminus \{T_k\} \cup \{t_n\})\}$  を探索する。
- A2-3.** もし  $T'_k = T_k$  ならば  $h = h+1$  とし、さもなければ  $h = 0$  として  $\mathcal{T}_K = \mathcal{T}_K \setminus \{T_k\} \cup \{T'_k\}$  のように更新する。
- A2-4.** もし  $h = K$  ならば  $\mathcal{T}_K$  を出力して終了する。
- A2-5.** もし  $k = K$  ならば  $k = 1$ 、さもなければ  $k = k+1$  とし、A2-2 に戻る。

明らかに、このアルゴリズムの計算量は改善が終わらない

限り増え続けてしまうが、ある程度大規模な問題に対しても、せいぜい貪欲法アルゴリズムの計算量  $O(NK)$  の数倍程度で終了することを我々は既に実験によって示している [8]。

### 2.2.3 提案解法

もし、計算量を最低限に抑えることを目的として、単純に貪欲法アルゴリズムと局所探索法アルゴリズムを組み合わせると、

- C1.** A1 で  $\mathcal{T}_K$  を得る。
- C2.** A2 で  $\mathcal{T}_K$  を改善する。

となる。確かに、これだけでも十分な近似解が期待できるが、スイッチングタイムステップ数  $K$  が貪欲法アルゴリズムによって決定されてしまうため、問題に対して不適切なスイッチングタイムステップ数のまま局所改善を行ってしまう恐れが大いにある。したがって我々は、不必要なスイッチングタイムステップは極力追加せず、且つ必要なスイッチングタイムステップは極力追加することを目的とした、アルゴリズムの反復的な組み合わせを提案する。提案解法の手順は以下となる。

- P1.** A1-1 から処理を開始する。
- P2.** A1-4 の処理後に  $k \geq 2$  ならば、 $\mathcal{T}_k$  を  $\mathcal{T}_K$  として出力する。
- P3.**  $\mathcal{T}_K$  を A2 で改善し、改善した  $\mathcal{T}_K$  を  $\mathcal{T}_k$  として出力する。
- P4.** A1-5 から処理を再開させ、ステップ I2 へ戻る。

この手順では、スイッチングタイムステップが追加される度に局所探索法アルゴリズムを行うため、更なる計算量の増加が予想されるが、ある程度大規模な問題に対しても、せいぜい貪欲法アルゴリズムの計算量  $O(NK)$  の数倍から十数倍程度で終了することを我々は既に実験によって示している [8]。

上記の解法によって得られた推定タイムステップ集合を  $\hat{\mathcal{T}}_K$  とし、各カテゴリ  $j$  について、タイムステップ  $n \in \mathcal{N}_k$  ( $0 \leq k \leq K$ ) における確率関数を  $\hat{p}_j(n) = \hat{p}_{k,j}$  のように考える。なお、上記の解法は、多項分布のレジームスイッチングを想定した人工データにおける実験で、極端に短いレジームの場合を除いて、真の分布に基いてパラメータを設定した Kleinberg の手法 [1] と同等、もしくはそれ以上の検出精度を示している [9]。

## 3. 実験結果

実験で用いる現実データは、goo 天気 \*1 のデータである。今回、全国 56 箇所の地上気象観測所における 1961 年から 2016 年の天気情報を対象データとした (那覇, 石垣

\*1 <https://weather.goo.ne.jp/>

島, 宮古島, 南大東島の観測所は1964年から2016年). 実験時には各観測所の1年ごとのデータを  $D$  として提案手法を適応しているが, 観測所や年によって日ごとの観測回数が異なるため, 観測数  $|D| = N$  はデータごとに異なることに注意されたい. なお, 各観測所において出現確率が1%未満の天気状態は欠損扱いとしており, カテゴリ  $J$  には含まれていない.

まず, 図1に検出されたスイッチングタイムステップ数  $K$  の度数分布を示す. 図より, ほとんどのデータは3個から8個の確率分布として表現されていることがわかる. なお,  $K$  の平均は6.0179である. 次に, 観測ステップ数  $N$

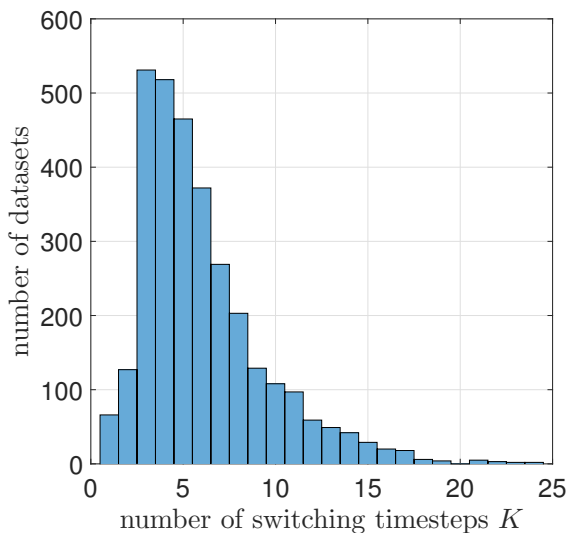


図1 スイッチングタイムステップ数  $K$  の度数分布

に対してスイッチングタイムステップ数  $K$  が多い観測所を調べる. 表1に示すように, 今回のデータは天気状態数  $J$  によって大きく2グループに分けられるため, 天気状態数  $J=3$  (晴れ, 曇, 雨) のグループと天気状態数  $J=4$  (晴れ, 曇, 雨, 雪) のグループに分けて集計する. なお, 天気状態数  $J=5$  (晴れ, 曇, 雨, 雪, 霧) の観測所は釧路と室蘭である.  $J=3$  (晴れ, 曇, 雨) のグループの

$J$	observatories	rate
3	28	50.00%
4	26	46.43%
5	2	3.57%

表1 天気状態数  $J$  ごとの観測所数

観測ステップ数  $N$  に対するスイッチングタイムステップ数  $K$  を表2に示す. 表より, 上位には沖縄本島と離島が, 下位には瀬戸内海に面した地域が集中していることがわかる. 提案モデルに基づいた解釈としては,  $N$  の総和に対して  $K$  の総和が大きい (小さい) ほど, その地域の天気傾向は時期と共に変わりやすい (変わりにくい) と言える.

同様に,  $J=4$  (晴れ, 曇, 雨, 雪) のグループの観測ス

rank	observatory	sum of $K$	sum of $N$	rate of $K$
1	石垣島	629	58961	0.0107
2	宮古島	587	56619	0.0104
3	宮崎	628	62247	0.0101
4	南大東島	573	56837	0.0101
5	那覇	584	58962	0.0099
6	前橋	536	62249	0.0086
7	東京	521	62207	0.0084
8	銚子	511	62249	0.0082
9	甲府	504	62249	0.0081
10	鹿児島	503	62193	0.0081
11	名古屋	481	62249	0.0077
12	大分	476	62249	0.0076
13	熊谷	311	41527	0.0075
14	熊本	449	62219	0.0072
15	宇都宮	292	41527	0.0070
16	静岡	290	41538	0.0070
17	水戸	288	41527	0.0069
18	高松	418	62209	0.0067
19	高知	278	41544	0.0067
20	横浜	274	41532	0.0066
21	大阪	406	62249	0.0065
22	津	252	41442	0.0061
23	徳島	251	41575	0.0060
24	奈良	242	41529	0.0058
25	岡山	232	41535	0.0056
26	松山	346	62249	0.0056
27	神戸	229	41511	0.0055
28	和歌山	226	41542	0.0054

表2  $J=3$  (晴れ, 曇, 雨) のグループの観測ステップ数  $N$  に対するスイッチングタイムステップ数  $K$

テップ数  $N$  に対するスイッチングタイムステップ数  $K$  を表3に示す. 上位の地域や下位の地域の明確な共通点はないが, 天気傾向について先程と同様の解釈ができる.

さらに, 提案手法の出力結果について詳しく見ていく.  $J=3$  (晴れ, 曇, 雨) グループで  $N$  の総和に対して  $K$  の総和が最も高かった石垣島の, 年ごとのスイッチングタイムステップ数  $K$  と年ごとの対数尤度  $L(D; \hat{P}_K, \hat{T}_K)$  を図2, 3にそれぞれ示す. 両図より,  $K$  の数と  $L(D; \hat{P}_K, \hat{T}_K)$  の値はおおむね同じような変動となっているが, 必ずしも  $K$  の数が多い (少ない) ほど  $L(D; \hat{P}_K, \hat{T}_K)$  の値が大きく (小さく) なっているわけではないことに注意されたい. また,  $L(D; \hat{P}_K, \hat{T}_K)$  が最大となった2002年の出力結果は図4のようになっており, 10観測ステップごとに集計して時系列データ化した元のデータ (図5) と比較すると, 複雑な天気傾向の変化をシンプルに表現できていることが見て取れる.

最後に, 1964年から2016年までの出力結果から日ごと

rank	observatory	sum of $K$	sum of $N$	rate of $K$
1	稚内	391	62155	0.0063
2	福島	225	41479	0.0054
3	長崎	336	62249	0.0054
4	富山	223	41523	0.0054
5	山形	223	41529	0.0054
6	福井	222	41523	0.0053
7	佐賀	221	41514	0.0053
8	青森	330	62249	0.0053
9	長野	214	41509	0.0052
10	網走	319	62180	0.0051
11	旭川	312	62248	0.0050
12	函館	311	62204	0.0050
13	金沢	310	62176	0.0050
14	秋田	310	62248	0.0050
15	下関	305	62197	0.0049
16	札幌	301	62247	0.0048
17	松江	276	57425	0.0048
18	仙台	294	62249	0.0047
19	彦根	191	41526	0.0046
20	京都	189	41466	0.0046
21	盛岡	283	62248	0.0045
22	福岡	283	62249	0.0045
23	広島	282	62247	0.0045
24	鳥取	256	57352	0.0045
25	新潟	263	62210	0.0042
26	岐阜	173	41532	0.0042

表 3  $J = 4$  (晴れ, 曇, 雨, 雪) のグループの観測ステップ数  $N$  に対するスイッチングタイムステップ数  $K$

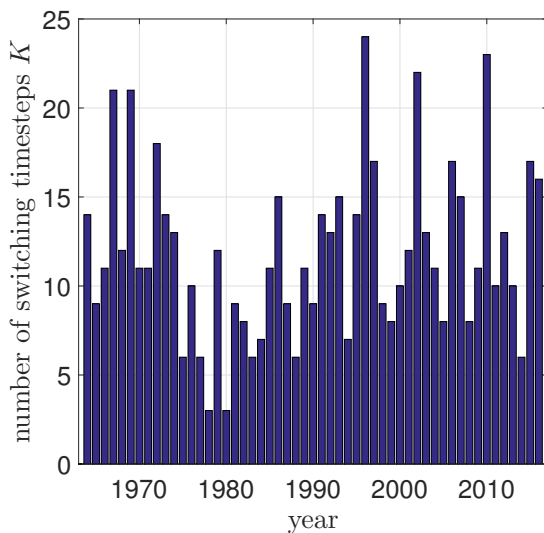


図 2 年ごとのスイッチングタイムステップ数  $K$  (石垣島)

の天気確率を集計し, 元のデータと比較する. 提案手法の出力結果から得られた日ごとの天気確率と元のデータから得られた日ごとの天気確率を図 6, 7 にそれぞれ示す. 両図より, 提案手法の出力結果は, 元のデータの特徴を最低

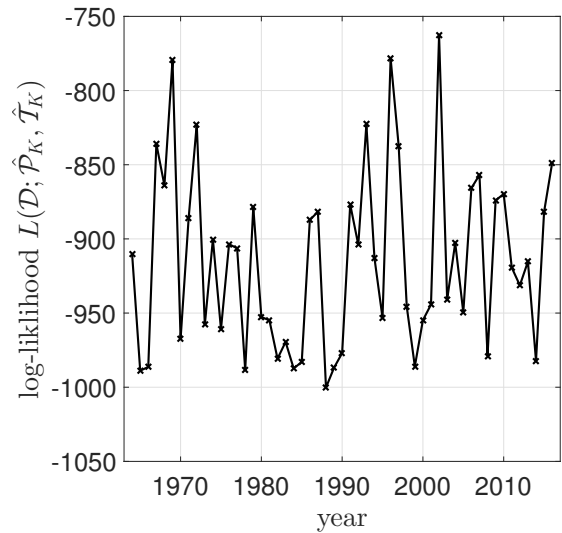


図 3 年ごとの対数尤度  $L(D; \hat{P}_K, \hat{T}_K)$  (石垣島)

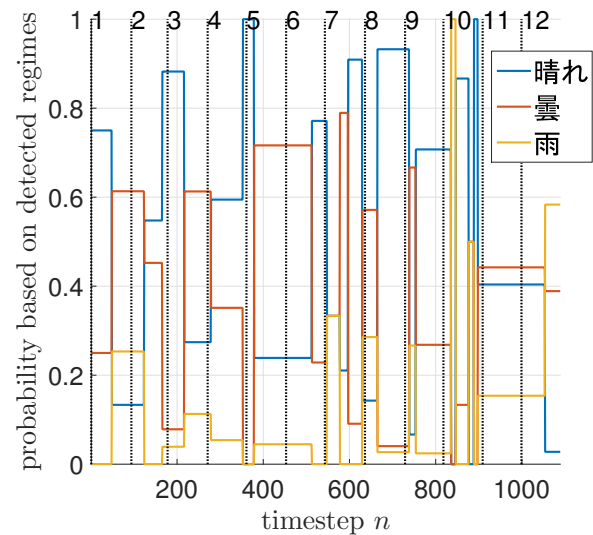


図 4 提案手法の出力結果 (石垣島 2012 年)

限の情報で再現できていることがわかる.

#### 4. おわりに

環境情報の圧縮や単純化を目的として, 多項分布に基づくレジームスイッチング検出法を提案し, 現実データを用いて提案手法の性能を評価した. 提案手法が生成したタイムラインは, 非常に単純な情報でまとめられているが, 元データの本質は失っていないことを実験で示した. 今後は, 多様な環境情報で実験を行い, 提案手法の有用性を検証する予定である.

#### 謝辞

本研究は, JSPS 特別研究員奨励費 16J11909 の支援を受けて行ったものである.

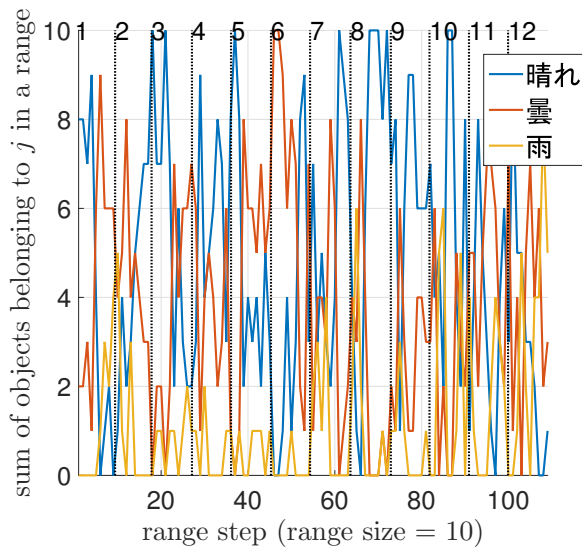


図 5 10 観測ステップごとに集計した天気 (石垣島 2002 年)

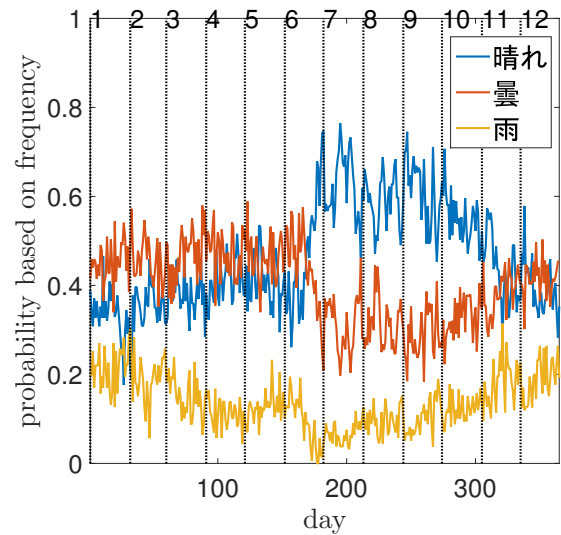


図 7 元のデータから得られた日ごとの天気確率 (石垣島)

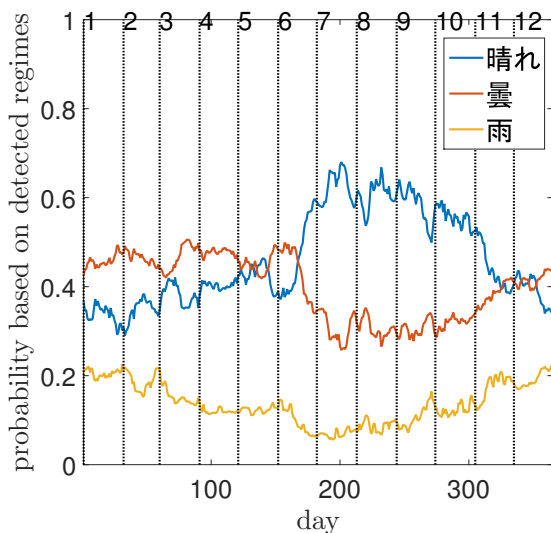


図 6 提案手法の出力結果から得られた日ごとの天気確率 (石垣島)

#### 参考文献

- [1] Kleinberg, J.: Bursty and hierarchical structure in streams, *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002)*, pp. 91–101 (2002).
- [2] Swan, R. and Allan, J.: Automatic generation of overview timelines, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000)*, pp. 49–56 (2000).
- [3] Zhu, Y. and Shasha, D.: Efficient Elastic Burst Detection in Data Streams, *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)*, pp. 336–345 (2003).
- [4] Sun, A., Zeng, D. and Chen, H.: Burst Detection from Multiple Data Streams: A Network-based Approach, *IEEE Transactions on Systems, Man, & Cybernetics Society, Part C*, Vol. 40, pp. 258–267 (2010).
- [5] Chandola, V., Banerjee, A. and Kumar, V.: Anomaly Detection: A Survey, *ACM Comput. Surv.*, Vol. 41, No. 3,

pp. 15:1–15:58 (2009).

- [6] Kim, C. J., Piger, J. and Startz, R.: Estimation of Markov regime-switching regression models with endogenous switching, *Journal of Econometrics*, Vol. 143, pp. 263–273 (2008).
- [7] Josang, A., Ismail, R. and Boyd, C.: A survey of trust and reputation systems for online service provision, *Decision support systems*, Vol. 43, pp. 618–644 (2007).
- [8] Yamagishi, Y., Okubo, S., Saito, K., Ohara, K., Kimura, M. and Motoda, H.: A Method to Divide Stream Data of Scores over Review Sites, *Proceedings of the 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI '14)*, pp. 791–800 (2014).
- [9] Yamagishi, Y. and Saito, K.: Visualizing Switching Regimes Based on Multinomial Distribution in Buzz Marketing Sites, *Proceedings of the 23rd International Symposium on Methodologies for Intelligent Systems (ISMIS '17)*, pp. 385–395 (2017).