物理量の時空間分布推定と時空間特異点検知

Time and Spatial Distribution Estimation of Physical Parameters and Singularity Detection

武内 俊樹 †　　岡留 剛 †
Toshiki Takeuchi　Takeshi Okadome

概 要

西日本に起こった平成 30 年 7 月豪雨のような 100 年に 1 度と言われる異常気象（特異点）に対して，事前に予測することは一般に難しい．特異点とは，過去に経験した現象や観測したデータと比べ，大きく外れた現象・データのことを示す．ただし，毎年のように起こる現象はいかなるものでも特異点ではなく，正常なものとして扱われることに注意する．本研究では，気象において過去に前例のない（100 年に 1 度起こると言われるような）特異点の検知を目的とし，物理量の時空間分布推定と時空間特異点検知手法を提案する． 時空間分布推定では，高精度センサと低精度センサを利用し，個々のセンサ値を補正または欠損値を推定する．時空間分布推定により，高精度センサがない地域においても特異点検知手法を適用できる利点がある．特異点検知手法では，ある地域の降水量時系列データを部分時系列に変換し，部分時系列における各ベクトルを多次元空間上の 1 点に対応づける．多次元空間上で近傍距離を計算し，一定の基準距離（閾値）と比較することで部分時系列ごとに異常を判定する．予備実験として，降水量において過去に特異点があった 20 箇所の地域に対し，提案手法によって対象の特異点を早く確実に捉えられるか検証した．

For abnormal weather (singularity) said to be once in 100 years like heavy rain in July 2018 which happened to western Japan, It is generally difficult to predict in advance. Singularity refers to phenomena and data that deviate significantly from phenomena experienced in the past and observed data. However, it should be noted that any phenomena that occur every year are treated as normal ones, not as singular points. In this study, we propose time and spatial distribution estimation of physical parameters and singularity detection methods for the purpose of detecting unprecedented singularity in the weather (which is said to occur once in 100 years). In time and spatial distribution estimation, high accuracy sensors and low accuracy sensors are used to correct individual sensor values or estimate missing values. By time and spatial distribution estimation, there is an advantage that the singular point detection method can be applied even in areas without high precision sensors. In the singularity detection method, the precipitation time series data of a certain area is converted to a partial time series, and each vector in the partial time series is associated with one point in the multidimensional space. The neighborhood distance is calculated in multidimensional space, and the anomaly is judged for each partial time series by comparing the fixed reference distance (threshold). As a preliminary experiment, we verified whether the proposed method could capture the singularity of the object quickly and reliably by the proposed method, for the 20 regions that had singularity in the past in the precipitation amount.

1. はじめに

西日本に起こった平成 30 年 7 月豪雨のような 100 年に 1 度と言われる異常気象（特異点）に対して，事前に予測し判断することは一般に難しい．特異点とは，過去に経験した現象や観測したデータと比べ，大きく外れた現象・データのことを示す．主に特異点検知手法は，機器の故障やビジネスにおいて売り上げ変化を捉えるといった実用的な分野で活用されている．記録的豪雨に代表される異常は過去に前例のない降水量を観測するため，機器の故障時に機器が過去にはない動作をする傾向と同じ側面がある．気象においても過去に前例のない異常は，従来の特異点検知手法を適用できる可能性がある．

そこで本研究では，気象において過去に前例のない（100 年に 1 度起こると言われるような）特異点の検知を目的とする．物理量は降水量を扱い，特異点は記録的豪雨などに着目する．アメダス雨量値（高精度）とレーダー雨量値（低精度）を利用する．ただし，毎年のように起こる降水パターンはいかなるものでも特異点ではなく，正常なものとして扱われることに注意する．また，観測する物理量における異常は時間変動として極少数であるものとする．本研究は，物理量の時空間分布推定と，時空間特異点検知の大きく 2 段階に分け，手法を提案する．

† 関西学院大学大学院，Kwansei Gakuin University Graduate

時空間分布推定では，高精度センサと低精度センサを利用し，個々のセンサ値を補正または欠損値を推定する．推定したい空間を特定のメッシュで区切り，メッシュの格子点における物理量を，マルコフ確率場を利用して空間分布推定を行なう．実用上，環境物理量には不連続性を考慮しないといけないため，ラインプロセス手法により不連続領域を検出する．時空間分布推定を行なうために，連鎖グラフを導入し確率変数間の因果関係を持たせることで時系列データの扱いを可能とする．レーダー雨量値は日本全国において降水量を観測したものであり，ノイズが大きく欠損値があるため，高精度センサがない地域で特異点検知手法が使用できない問題がある．時空間分布推定を行なうことで，アメダスがない地域に対しても特異点検知手法を適用できる利点がある．

特異点検知手法では，ある地域の降水量時系列データを部分時系列に変換し，部分時系列における各ベクトルを多次元空間上の 1 点に対応づける．多次元空間上で近傍距離を計算し，一定の基準距離（閾値）を比較することで部分時系列ごとに異常を判定する．予備実験として，降水量において過去に特異点があった 20 箇所の地域に対して，提案手法によって対象の特異点を早く確実に捉えられるか検証した．その結果，20 箇所で観測史上 1 位となる記録的豪雨を 24 時間以上前に検知でき，高い確率で特異点検知が可能であることを示した．しかし，降水量のみでは様々な特性を持つ特異点を捉える精度は低く，降水量以外の物理量を利用した手法の検討が今後の重要な課題である．本論文の貢献は，レーダー雨量の観測可能な地点ならば，特異点検出手法を気象に適用し，記録的豪雨といった特異点は検出可能としたことである．以下各節で，関連研究と提案手法・予備実験について詳細に述べまとめを記す．

2. 関連研究

車両機器に関する異常検知手法として，近藤稔 [1] の手法がある．機器の振動や音の音色をデータ化し，時系列データを部分時系列に変換した後，多次元空間上の 1 点の座標に対応付ける．対応付けられた点と他の点との近傍距離から異常度を算出し，ある閾値と比較することで，人が判断するよりも早く故障を発見することが可能である．本研究では，降水量データを用いて，人よりも早く特異点を発見することを目的とする．

また，機械学習と気象衛星ひまわり 8 号の雲画像による台風検出手法として，金崎拓郎ら [2] の手法がある．過去の気象衛星雲画像を大量に用いて台風位置を学習することで，発達前の台風や衰弱した台風を含めて検出できる．しかし，異常気象の中には突発的なものも含まれ，台風以外が原因で発生するものも存在する．本研究では，過去のデータに前例がないパターンを発見することに着目し，気象予報士でも予測困難である特異点を予測することを目的としている．

3. 時空間分布推定

本研究は，物理量の時空間分布推定と時空間特異点検知の 2 段階に分けて説明する．まずは時空間分布推定手法を説明し，後に時空間特異点検知手法を説明する．

図 1 に示すように，時空間分布推定では，地上レーダー雨量（低精度）と地域気象観測システムから得られるアメダス雨量（高精度）を用いる．地上レーダー雨量は，日本全国の雨量を観測することができるが，ノイズが大きく，時刻によっては欠損値が存在する．そのため，アメダスがない地点において特異点検知を行なう場合は，地上レーダー雨量のノイズを小さくし，欠損値を推定した後が望ましい．まず本研究では，少数の高精度センサと多数の低精度センサを用いてセンサフュージョンを行ない，物理量の時空間分布推定を行なう．さらに，ラインプロセス手法により局所的な降水量の変化を捉え，不連続性を考慮する．

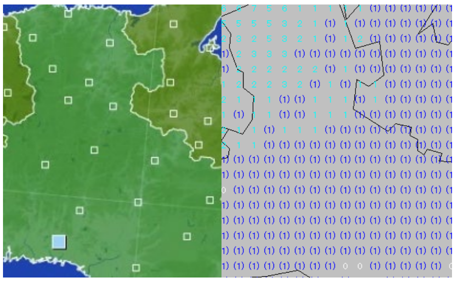

図 1: アメダス雨量（左）と地上レーダー雨量（右）の例．気象庁が提供する気象観測データより (Copyright(c) 2006 by Japan Meteorological Agency)．

3.1 物理量の不連続性を考慮した空間分布推定

まず時空間分布推定の前に，マルコフ確率場 [3] を利用した空間分布推定を説明する．マルコフ確率場（無向グラフ）により，確率変数間の緩い依存関係を表現した例を図 2 に示す．推定したい空間を決め，地上レーダー雨量の観測点を通るようにメッシュで切る．メッシュの格子点に真の値を反映すると思われる i 番目の確率変数を x_i とする．低精度センサの観測値は図中における水色のノードに対応し，高精度センサの観測値は赤色のノードに対応する．例えば，図 2 右のモデルにおける中央のノードは隣接する 4 方向のノードと、対応しているセンサに依存関係がある．また，無向エッジにはそれぞれ対応するエッジ変数があるとする．

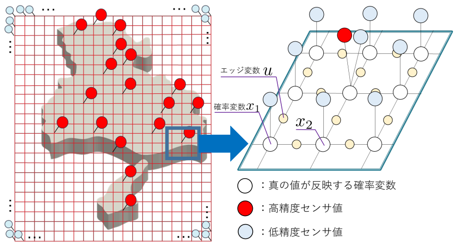

図 2: 兵庫県をメッシュで切った例（左）と 3×3 格子の例（右）．

このとき，マルコフ確率場の同時確率は，クリークの概念を導入すると便利に表現できる．クリークは全結合の条件を満たしたグラフの部分集合であり，あと 1 つノードを加えると全結合ではなくなる状態のクリークを極大クリークとする．クリークは C で表し，クリーク内の変数集合を $\mathbf{x}_C$ とし，クリーク内のエッジ変数集合を $\mathbf{u}_C$ とする．Z は規格化定数である．同時分布 $p(\mathbf{x},\mathbf{u})$ に対して，極大クリーク上のポテンシャル関数の積で表現したものを以下の式とする．

$$p(\mathbf{x},\mathbf{u}) = \frac{1}{Z}\prod_{C}\psi_C\left(\mathbf{x}_C,\mathbf{u}_C\right) \tag{1}$$

$$\psi_C\left(\mathbf{x}_C,\mathbf{u}_C\right) = \exp\left\{-E\left(\mathbf{x}_C,\mathbf{u}_C\right)\right\} \tag{2}$$

ポテンシャル関数はエネルギー関数 $E\left(\mathbf{x}_C,\mathbf{u}_C\right)$ によって決まるものとする．このエネルギー関数を最小化することによって，与えられた観測値にフィッティングする [4]．第 1 項と第 2 項はデータフィッティング項であり，第 1 項は低精度センサ，第 2 項は高精度センサと対応している．第 3 項は拘束条件項であり，隣接する確率変数間の依存関係を表現している．さらに，拘束条件項にラインプロセス手法 [5] を導入し，不連続を考慮可能としている．なお，ラインプロセス手法の詳細は次節説明する．格子点の総数を N とし，$\mathbf{x} = \{x_1, x_2, ..., x_N\}$ とする．D_{low} は低精度センサ値を持つ格子点の集合であり，i 番目の格子点に対応する低精度センサ値を l_i とする．D_{high} は高精度センサ値を持つ格子点の集合であり，i 番目の格子点に対応する高精度センサ値を h_i とする．ただし，高精度センサの数は低精度センサの数に比べ圧倒的少数とする．$bd(x_i)$ は確率変数 x_i に隣接する確率変数集合である．このとき，u_a は確率変数 x_i と確率変数 x_i に隣接する確率変数 a を接続するエッジに対応したエッジ変数とする．σ はシグモイド関数であり，式 (5) で表される．c はラインプロセス手法における閾値であり，不連続領域の検出しやすさを意味する．各項に関わるパラメータは，使用するセンサのダイナミックレンジで決定する．

$$\begin{aligned} E\left(\mathbf{x}_C,\mathbf{u}_C\right) =& w_0 \sum_{i\in D_{low}} \{l_i - x_i\}^2 \\ &+ w_1 \sum_{i\in D_{high}} \{h_i - x_i\}^2 \\ &+ w_2 \sum_{i\in N}\sum_{a\in bd(x_i)} f(x_i, a) \end{aligned} \tag{3}$$

$$\begin{aligned} f(x_i, a) =& \{x_i - a\}^2\{1-\sigma\left(u_a\right)\} \\ &+ c\sigma\left(u_a\right) \\ &+ \sigma\left(u_a\right)\{1-\sigma\ \left(u_a\right)\} \end{aligned} \tag{4}$$

$$\sigma = \frac{1}{1+e^{-u_a}} \tag{5}$$

関数 f は，ラインプロセスを含んだ関数であり式 (4) で表現される．以上の式を定義し，同時確率 $p(\mathbf{x},\mathbf{u})$ が観測値が与えられたもとで事後確率最大になるよう推定を行なう．実用上は物理量の不連続領域を考慮しなければならず，式 (3) の第 3 項がその役割を果たしている．次節でラインプロセス手法による不連続領域を考慮する手法を示す．

3.2　ラインプロセス手法による不連続領域検出

環境物理量の空間分布推定において，実用的には不連続領域の検出が必要である．不連続領域とは，照度においては人の影の領域，雨量においては浮遊物によって雨が遮断される領域，温度においては建築物の壁などがある．不連続領域における観測は周囲とは全く異なる値を取るため，センサフュージョンの際には精度低下の原因になる．そこでラインプロセス手法により，不連続領域を別の領域として分割し，領域ごとに空間分布推定を行なう．

ラインプロセス手法は，式 (4) の関数 f に既に反映されている．2 乗誤差は確率変数 x_i と x_i の隣接する確率変数 a との依存関係を表しており，x_i と a の値が近づくほど，エネルギー関数によって返される値が小さくなる．式 (4) は閾値 c よりも 2 乗誤差が小さい場合は二乗誤差によって得られる値をエネルギー関数に加え，大きい場合は閾値 c をエネルギー関数に加えることを意味する．なぜなら，エネルギー関数の最小化を行なうため，最小化に有利となる値を選択していくからである．すなわち，2 乗誤差と閾値 c を比較し，値が大きい方はエネルギー関数を大きくするために無視し，値が小さい方をエネルギー関数に加える．この流れは，ユーザーの手によらず式 (3) のエネルギー関数を最小化する過程で行なわれる．その仕組みを図 3 に示す．

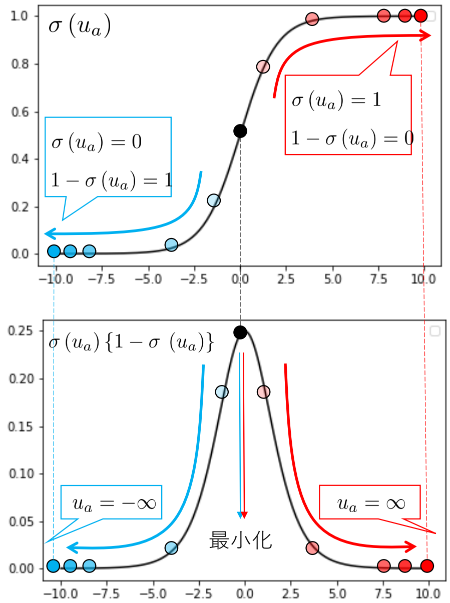

図 3: 関数 f が最小化で起こる動作の様子．横軸 u_a．

関数 f が最小化で起こる動作として，図 3 の下図から u_a が ∞ か $-\infty$ に徐々に近づいていくことを意味する．近づく速さが遅い場合は，重み付けなどによる工夫が必要である．u_a が ∞ に近づいた場合，関数 f において二乗誤差項が消えるので，エネルギー関数に含まれず依存関係がなくなる．その代わり閾値 c がエネルギー関数に加わるが，閾値は定数であるためエネルギー関数の最小化には何も影響を与えない．これが不連続領域と解釈される．逆に，u_a が $-\infty$ に近づいた場合，関数 f において閾値 c が消えることになり，二乗誤差項がエネルギー関数に加えれる．二乗誤差項の値が，閾値 c よりも小さい場合は連続領域と解釈される．

3.3　連鎖グラフを用いた時空間分布推定

これまで解説した空間分布推定から，時系列データを扱う時空間分布推定を考える．無向グラフは確率変数間の緩やかな依存関係を表し，有向グラフは確率変数間の因果関係を表す．時系列を扱うためには，時刻間の因果関係を表現する有向グラフが必要であるため，連鎖グラフ [6] [7] を導入する．連鎖グラフは，無向グラフと有向グラフを融合したグラフであり，有向エッジと無向エッジを両方持つグラフに拡張できる．

連鎖グラフを扱う上で，新たに「ブロック」と「順序」の 2 つの概念を導入する．ブロックとは，無向グラフのみで形成される部分グラフのことであり，ブロック内のグラフは有向エッジを持たないとする．例えば図 4 の左図において，ブロック $b_1=\{1,2\}$，$b_2=\{3\}$，$b_3=\{4,5\}$ の 3 つのブロックに分けられる．ブロックの概念を導入したことにより，ブロック同士は有向エッジのみで連結され，ブロック間に順序の概念が生まれる．連鎖グラフではこのブロックと順序を用いて時系列データを扱う．また，図 4 中央は有向グラフであり，ブロック $b_1=\{1\}$，$b_2=\{2\}$，$b_3=\{3\}$，$b_4=\{4\}$，$b_5=\{5\}$ のブロック集合を持つ連鎖グラフと言える．図 4 右は無向グラフであり，$b_1=\{1,2,3,4,5\}$ のブロック集合を持つ連鎖グラフと言えるため，連鎖グラフは無向グラフと有向グラフを内包する．

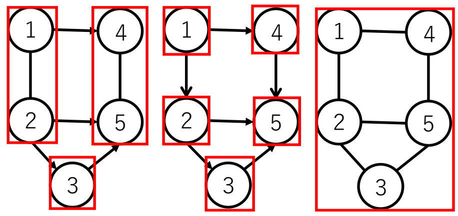

図 4: 連鎖グラフの例．

前章に紹介した空間分布推定の役割をする無向グラフに，t 時刻目の状態を表す潜在変数を $z^{(t)}$ を対応させる．図 5 に格子を 2 × 2，時系列を 2 とした提案モデルの連鎖グラフを示す．図 5 のブロックは，$b_1=\left\{z^{(1)}\right\}$，$b_2=\left\{x_1^{(1)},x_2^{(1)},x_3^{(1)},x_4^{(1)},l_1^{(1)},l_2^{(1)},h_1^{(1)}\right\}$，$b_3=\left\{z^{(2)}\right\}$，$b_1=\left\{x_1^{(2)},x_2^{(2)},x_3^{(2)},x_4^{(2)},l_1^{(2)},l_2^{(2)},,h_1^{(2)}\right\}$ となる．潜在変数間に連結される有向エッジは時刻間の因果関係を表し，潜在変数 z と確率変数 x を連結する有向エッジは，状態と真の値と思われる確率変数との因果関係を表す．

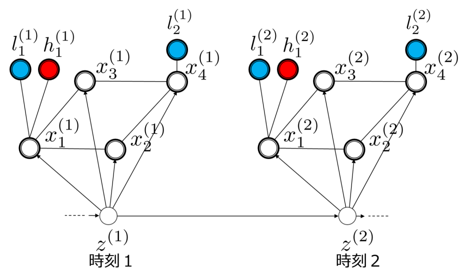

図 5: 2 × 2 格子，時系列の長さ 2 の連鎖グラフ．

低精度センサ値集合 $\mathbf{l}$，高精度センサ値集合 $\mathbf{h}$，学習済みパラメータ集合 θ が与えられたもとで，真値が反映される確率変数集合 $\mathbf{x}$，潜在変数集合 $\mathbf{z}$，エッジ変数集合 $\mathbf{u}$ の推定を行なう．$\theta=\{\pi,\mathbf{A},\phi\}$ であり，遷移確率 $\mathbf{A}$，出力確率 ϕ，初期要素は π とする．潜在変数の状態数は訓練データを分割し，最も精度の良い状態数を与えるものとする．無向エッジのみで接続されたノード集合であるブロックの概念を導入し，グラフのブロック数を B としたとき，連鎖グラフの同時分布は，ブロックの積の形として以下の式で因数分解の表現ができる．

$$p(\mathbf{x},\mathbf{z},\mathbf{u}|\mathbf{l},\mathbf{h},\theta)=p(b_1)\prod_{i=1}^{B}p(b_i|b_{i-1}) \qquad (6)$$

ブロック単体は無向グラフであるため，ブロックごとにマルコフ確率場の枠組みを与える．ブロックごとに同時分布を計算することで確率的解釈が可能となり，$p(\mathbf{x},\mathbf{z},\mathbf{u})$ の同時分布が最大になるように推定を行う．なお図5において，大域的マルコフ性 [7] により，連鎖グラフにおける条件付き独立性 $b_2 \perp\!\!\!\perp b_3 \,|\, b1$ があるためブロックの積で表現できる．例として，図5の2時刻における連鎖グラフの因数分解を考えると，以下の式になる．

$$p(b_1,b_2,b_3,b_4) = p(b_1)p(b_2\,|\,b_1)p(b_3\,|\,b_1)P(b_4\,|\,b_3) \quad (7)$$

$$p(b_1) = p(z^{(1)}) \quad (8)$$

$$p(b_2\,|\,b_1) = p(x_1^{(1)},x_2^{(1)},x_3^{(1)},x_4^{(1)},l_1^{(1)},l_2^{(1)},h_1^{(1)}\,|\,z^{(1)}) \quad (9)$$

$$p(b_3\,|\,b_1) = p(z^{(2)}\,|\,z^{(1)}) \quad (10)$$

$$p(b_4\,|\,b_3) = p(x_1^{(2)},x_2^{(2)},x_3^{(2)},x_4^{(2)},l_1^{(2)},l_2^{(2)},h_1^{(2)}\,|\,z^{(2)}) \quad (11)$$

このとき，ある時刻 t における $p(b_{i+1}\,|\,b_i)$ は前節の説明と同様，クリークの概念を用いて因数分解を行なう．なお，連鎖グラフのモラルグラフとして，有向エッジは無向エッジへと変換している．

$$\begin{aligned}p(b_{i+1}\,|\,b_i) = &\frac{1}{Z_i^{(t)}}\psi_1(x_1^{(t)},x_2^{(t)})\\ &\psi_1(x_1^{(t)},x_3^{(t)})\psi_1(x_2^{(t)},x_4^{(t)})\\ &\psi_1(x_3^{(t)},x_4^{(t)})\psi_1(x_1^{(t)},l_1^{(t)})\\ &\psi_1(x_1^{(t)},h_1^{(t)})\psi_1(x_4^{(t)},l_2^{(t)})\\ &\psi_2(z^{(t)},x_1^{(t)})\psi_2(z^{(t)},x_2^{(t)})\\ &\psi_2(z^{(t)},x_3^{(t)})\psi_2(z^{(t)},x_4^{(t)})\end{aligned} \quad (12)$$

ポテンシャル関数 ψ_1 を決定するエネルギー関数は前章で設定した式 (3) を利用する．ただし，状態の潜在変数を含むポテンシャル関数 ψ_2 はガウス分布の式を利用する．レーダー雨量値における欠損値などを推定することが可能となり，アメダスがない地域において時空間特異点検知が可能となる．

4. 時空間特異点検知

時系列データを任意の長さで分割した部分時系列ごとに異常を判定する．過去の前例のない異常を捉えることが目的であり，例年のように頻発している事象は特異点には含めないものとする．手法の流れとしては，まず部分時系列によって時系列データを任意の部分時系列データに変換し，多次元空間上に1点に対応付ける．対応付けられた点に対して，近傍距離を利用し異常判定するために参考する異常度を定義する．異常を判定する閾値として基準距離を決定し，基準距離より大きい場合は対応する部分時系列データは異常であると判定する．最後に，気象庁が提供する降水量データを用いて，過去の特異点を提案手法で検出可能か予備実験を行なった．

4.1 時系列データを部分時系列に変換

降水量の時系列データの中から異常部位を検出するため，1つの時系列データをスライド窓によって複数の部分時系列に表現する [8] [9]．スライド窓の窓幅を3時間に設定した場合，3時間ごとの部分時系列に分割される．例として，i 番目の地点における長さ T の時系列データを以下の式で表現する．

$$\mathcal{D} = \left\{x_i^{(1)},x_i^{(2)},\ldots,x_i^{(T)}\right\} \quad (13)$$

式 (13) を M 次元の部分時系列に変換する際，式 (13) の時系列データは以下の式で変換される．

$$\mathcal{D} = \left\{\mathbf{m}_i^{(1)},\mathbf{m}_i^{(2)},\ldots,\mathbf{m}_i^{(N)}\right\} \quad (14)$$

$$N = T - M + 1 \quad (15)$$

このとき，$\mathbf{m}_i^{(t)}$ は M 次元ベクトルであり，$\mathbf{m}_i^{(t)} = \left\{x_i^1,x_i^2,...,x_i^M\right\}$，$\mathbf{m}_i^{(t+1)} = \left\{x_i^2,x_i^3,...,x_i^{M+1}\right\}$ となる．本研究で扱う降水量の時系列データを部分時系列に変換する様子を図6に示す．図中のグラフは，岡山県倉敷市の2018年1月1日1時から2018年12月31日24時までにおける1時間ごとの降水量の推移を緑線で表している．なお，部分時系列は1時刻ずつスライドして分割していく．

部分時系列ごとに，異常度（大きいほど異常であることを示す）を求め，異常部位がどこか判定する．部分時系列ごとに異常度を計算することで，継続した異常を検知でき，観測した際のノイズによる誤検知を減らすことができる．異常度の計算と，異常部位の判定方法は次の章で詳細を記す．窓幅が小さいほど集中豪雨などの短期的な異常気象を捉えることに便利であり，窓幅が大きいと梅雨前線による多雨など長期的な異常気象を捉えることに便利である．

4.2 近傍法による異常度の決定

部分時系列の変換によって得られたデータ $\mathcal{D}$ を訓練データとし，新たに与えられたデータが異常か判定する．このとき部分時系列の各要素を，多次元空間上の1点の座標に対応付ける．2018年6月と7月における倉敷市の降水量の推移を，スライド窓の窓幅は2として，多次元

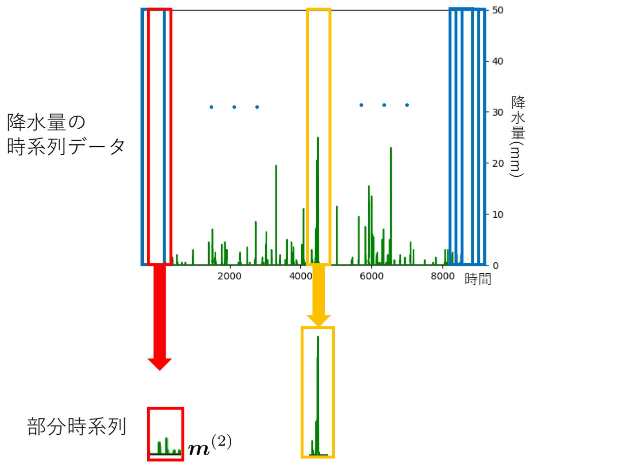

図 6: 時系列データを部分時系列に変換する図．アメダス雨量値は気象庁ホームページ（https://www.data.jma.go.jp/gmd/risk/obsdl/index.php）より参照．

空間に対応づける様子を図に示す．窓幅は 2 であるため，2 次元空間に対応付けられる．図に示した 2 次元空間の横軸は，2 時刻で分割されるうちの 1 つ目の時刻における値を表しており，縦軸は 2 つ目の時刻における値を表す．青で示した 2 時刻は，多次元空間に対応づけると，1 箇所に固まることがわかる．赤で示した 2018 年 7 月豪雨における 2 時刻を多次元に対応づけると，他のデータと離れた位置にプロットされる．ただし，訓練データは正常標本がほとんどであり，異常標本は限りなく少ないものとする．

異常度を k 近傍法によって，他のデータ間とのユークリッド距離で決定する．新しく与えられるデータベクトル $\mathbf{m}$ の異常度を，$a(\mathbf{x})$ としたとき，異常度は以下の式で近藤稔 [1] より定義されている．

$$a(\mathbf{x}) = \frac{1}{k_{NN}} \sum_{k=1}^{k_{NN}} \frac{|\mathbf{x} - NN_k(\mathbf{x})|}{d_k} - 1 \qquad (16)$$

k_{NN} は近傍データ数を表し，NN_k は入力データ $\mathbf{m}$ と k 番目に近いデータを表す．d_k は k 番目に近いデータに対応する基準距離（閾値）であり，基準距離は k 番目ごとに別々に値を求める必要がある．入力データ $\mathbf{m}$ と他のデータ点との距離を基準距離で割り 1 を引くと，負のときは入力データ $\mathbf{m}$ は正常であると判断し，正のときは入力データ $\mathbf{m}$ は異常であることを表現する．異常度 $a(\mathbf{x})$ は，小さい値であるほど正常であり，大きい値であるほど異常である．1 近傍のみだと訓練データに偶然似た異常度があった場合，入力データとの距離が近く正常と判断される可能性がある．また時刻が隣り合うデータが両方とも異常であった場合，先の時刻は異常と判断されるが後の時刻は正常と判断される可能性があるため，複数

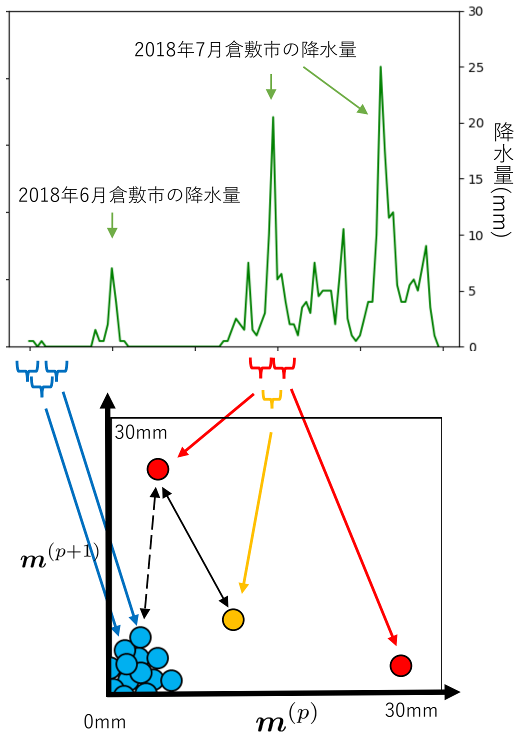

図 7: 部分時系列を多次元空間上に対応づける様子．

の近傍を考慮した方が精度が向上する．本研究では，訓練データを複数に分割し，最も精度が良かった近傍データ数 k_{NN} に決定する．

異常度の式 (16) で，図 7 における倉敷市の降水量データを，異常度に変換した様子を図 8 に示す．緑線は降水量の推移を表し，青線は異常度の推移を表している．異常度は 0 から 1 の間に収まるよう正規化しており，大きいほど異常であることを意味する．スライド窓の窓幅を 10 時間と設定している．スライド窓の幅を調節することで，捉えたい特異点の性質が異なる．本研究では，複数の窓幅を用いて，別々に異常度を計算する．

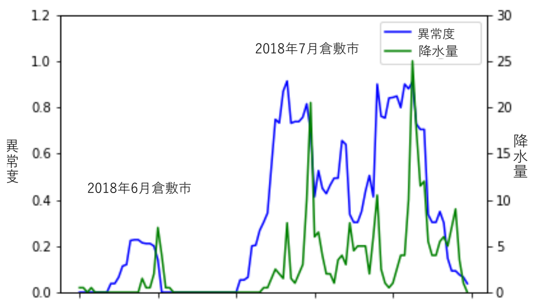

図 8: 図 7 の降水量データを異常度に変換した図．スライド窓の窓幅を 10 時間として異常度を計算している．アメダス雨量値は気象庁ホームページ（https://www.data.jma.go.jp/gmd/risk/obsdl/index.php）より参照．

4.3　基準距離の決定

新しい入力データが異常か正常かの判断のためには，基準距離を決定する必要がある．基準距離を大きくすると，正常標本を間違えて異常であると検出しないようになるが，逆に異常を見逃す問題がある．また，基準距離を小さくすると，確実に異常を発見するが，逆に正常標本を間違えて異常と判断する問題がある．そこで基準距離の決定には，この 2 つのトレードオフを考えるため，正常標本精度 r_0 と異常標本精度 r_1 を用いる．井出剛ら [8] は以下のように説明している．正常標本精度とは，実際に正常である標本の中で正しく正常と判定できた数を実際に正常である標本の総数で割った値である．異常標本精度とは，実際に異常である標本の中で正しく異常と判定できた数を実際に異常である標本の総数で割った値である．異常度の基準距離により，正常標本精度と異常標本精度は大きく変化するため，正常標本精度 r_0 と異常標本精度 r_1 を基準距離の関数とする．

正常標本精度と異常標本精度で異常の見逃しの指標としている．正常標本精度 $r_0 = 1$ のとき，正常標本に対しては精度が良く判定できるが，異常標本を発見できないことを意味する．異常標本精度 $r_1 = 1$ のとき，異常標本に対しては精度が良く判定できるが，正常標本を異常と判定することになる．正常標本精度と異常標本精度の 2 つの観点から見て，丁度良い箇所を分岐点とし，基準距離に設定する．実際には，正常標本精度と異常標本精度をグラフにプロットし，交差する点を調べる必要がある．

$$f \equiv \frac{2r_0 r_1}{r_0 + r_1} \tag{17}$$

実用上は式 (17) のように F 値を求めると便利である．異なる基準距離で繰り返し計算し，F 値が最大となる基準距離を探すことになる．近傍距離より異常度を定義しているため，各近傍ごとに基準距離を求める．

5.　予備実験

提案した時空間特異点検知手法を用いて，特異点を検出できるか実験を行なった．気象庁が提供するアメダス雨量値を利用する．対象とする特異点は，気象庁もしくは地方自治体の報告レポートなどに，異常気象であり，過去に事例がない事象かどうかはっきりと明記されているものとする．主に気象庁ホームページで公開されている「異常気象の特徴と要因に関する情報」の報告レポートを参照した．報告レポートで最大降水量の観測時刻の記載がある地点に着目した．またある一定期間の雨量における総量で，降水量の値を見ていくものとする．一定期間を 3 時間とした場合，3 時間の降水量の総量が観測史上最大であることを最大 3 時間降水量という．本実験では報告レポートに記載がある最大 3 時間降水量から最大 72 時間降水量まで扱う．気象庁の報告レポートに記載があった 20 箇所において特異点検出可能か，また最大降水量に至るどのくらい前に検出できるか検証した．以下表 1 に，今回の実験の対象となった特異点を記す．

表 1: 実験対象の特異点．出典：気象庁ホームページ（https://www.data.jma.go.jp/gmd/risk/obsdl/index.php）をもとに作成．地点名とは，アメダス雨量値を観測可能である箇所を示す．

	特異点	都道府県	市町村	地点名	降水量の時間幅
No.1		高知県	宿毛市	宿毛	6時間降水量
No.2		広島県	呉市	倉橋	12時間降水量
No.3		岡山県	笠岡市	笠岡	24時間降水量
No.4		兵庫県	養父市	八鹿	48時間降水量
No.5	平成30年7月豪雨	愛媛県	今治市	今治	72時間降水量
No.6		岐阜県	大野郡白川村	御母衣	72時間降水量
No.7		山口県	岩国市	岩国	72時間降水量
No.8		大阪府	豊中市	豊中	72時間降水量
No.9		鳥取県	八頭群若桜町	若桜	72時間降水量
No.10		福岡県	田川郡添田町	添田	72時間降水量
No.11		茨城県	古河市	古河	3時間降水量
No.12		宮城県	栗原市	鶯沢	24時間降水量
No.13	平成27年台風18号	埼玉県	越谷市	越谷	24時間降水量
No.14		栃木県	日光市	今市	24時間降水量
No.15		福島県	南会津群南会津町	舘岩	24時間降水量
No.16		三重県	津市	笠取山	3時間降水量
No.17		北海道	下川町	下川	3時間降水量
No.18	平成26年8月豪雨	京都府	福知山市	福知山	24時間降水量
No.19		石川県	羽咋市	羽咋	24時間降水量
No.20		徳島県	阿南市	蒲生田	48時間降水量

2008 年から 2018 年の 10 年間の降水量（1 時間ごと）を利用し，降水量の時系列データを部分時系列に変換し多次元空間上の 1 点に対応づける．式（16）より異常度を計算し，異常度が基準距離より大きい場合は特異点と判断する．スライド窓の窓幅は，3 時間，6 時間，12 時間，24 時間，48 時間，72 時間の窓幅を設定し実験を行なった．

5.1　実験結果

実験結果を表 2 に示す．表 2 の No.1 の記載は，表 1 の No.1 と対応している．最大降水量の観測時刻とは，対応する地点において観測史上 1 位となる降水量を観測した時刻である．なお一定期間の降水量の総量であり，表 1 の降水量の時間幅を参考にしている．特異点の検知時刻とは，特異点検知手法によって検知した時刻である．表 2 に示した時刻はスライド窓の窓幅を考慮した時刻であり，最も早く検知可能な時刻である．結果として，20 箇所のうち全ての箇所において異常検知可能であった．また，最大降水量の観測時刻より，特異点検知した時刻の

方が3日ほど早い結果となった．異常気象で大きな被害をもたらすほどの過去に前例のない豪雨が来る3日ほど前には，検知可能であることがわかる．

表 2: 実験結果を示す．最大降水量の観測時刻とは，対応する地点において観測史上 1 位となる降水量を観測した時刻である．特異点の検知時刻とは，特異点検知手法によって検知した時刻である．最大降水量の観測時刻より，特異点検知した時刻の方が早いことがわかる．出典:気象庁ホームページ（https://www.data.jma.go.jp/gmd/risk/obsdl/index.php）をもとに作成．

	最大降水量の観測時刻		特異点の検知時刻	
	月日	時分	月日	時分
No.1	2018/7/8	8:20	2018/7/5	7:00
No.2	2018/7/7	5:30	2018/7/4	7:00
No.3	2018/7/7	8:40	2018/7/3	13:00
No.4	2018/7/7	9:50	2018/7/4	10:00
No.5	2018/7/8	8:00	2018/7/5	7:00
No.6	2018/7/8	3:40	2018/7/5	3:00
No.7	2018/7/8	8:50	2018/7/2	10:00
No.8	2018/7/8	0:10	2018/7/5	0:00
No.9	2018/7/8	6:20	2018/7/2	5:00
No.10	2018/7/8	7:20	2018/7/2	6:00
No.11	2015/9/9	17:20	2015/9/6	14:00
No.12	2015/9/11	8:50	2015/9/8	7:00
No.13	2015/9/10	4:50	2015/9/7	7:00
No.14	2015/9/10	6:20	2015/9/6	13:00
No.15	2015/9/10	6:40	2015/9/7	5:00
No.16	2014/8/9	15:50	2014/8/6	17:00
No.17	2014/8/5	10:50	2014/8/2	13:00
No.18	2014/8/17	5:50	2014/8/13	22:00
No.19	2014/8/17	5:50	2014/8/14	6:00
No.20	2014/8/4	4:50	2014/7/31	5:00

気象庁から得られる降水量データは1時間ごとであるため，特異点検知時刻も1時間ごとの単位となる．問題点として，降水量を基準とした異常検知手法であるため，記録的豪雨などの異常気象しか捉えられない．異常気象には気温など関係するものもあり，降水量の他の環境物理量を利用して特異点を検知する必要がある．

6. まとめ

本稿では，物理量の時空間分布推定と時空間特異点検知を提案した．予備実験において，提案手法を用いて特異点検知手法を適用した結果，気象においても特異点は検知可能であることがわかった．また20地点において，最大降水量を観測する前に検知することが可能である．ただし，降水量データのみに着目しており，温度や湿度など他の物理量を活用し，記録的豪雨といった特異点とはまた異なる特異点を捉えていく必要がある．本研究における時空間分布推定と時空間特異点検知を併用し，レーダー雨量で観測可能な場所は多種多様な特異点検知が可能になることを目指す．

参考文献

[1] 近藤稔. 振動のオクターブバンド分析を用いた異常検知法による車両機器の診断. 日本機械学会論文, Vol. 84, No. 862, 2018.

[2] 金崎拓郎, 筆保弘徳, 加瀬紘熙, 松岡大祐, 吉田龍二. 機械学習を用いた台風検出器の開発. 電子情報通信学会信学技報, Vol. 118, No. 197, AI2018-15, pp. 13-18, 2018.

[3] Bishop, C. M *Pattern Recognition and Machine Learning*, Springer (2006).

[4] Ito, K. and H. Hontani (2006). Analysis of calibration performance of networked sensors using measurements graphs. *SICE-ICASE International Joint Conference*, 1320-1325.

[5] Geman, S. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Readings in computer vision*, 564-584.

[6] Zhang, L. (2011). Probabilistic image modeling with an extended chain graph for human activity recognition and image segmentation. *IEEE Transactions on Image Processing*, 2401-2413.

[7] 宮川雅巳. グラフィカルモデリング. 朝倉書店（2011）

[8] 井出剛, 杉山翔. 機械学習プロフェッショナルシリーズ 異常検知と変化検知. 講談社（2015）

[9] Shieh, J and E. Keogh (2008) . iSAX: Indexing and Mining Terabyte Sized Time Series. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD 08, 623-631.