

構造類似度指標の統合による SOM 気圧配置分類法の一考察

Synoptic Pattern Classification via Integrated Structural Similarity Metrics in SOM

高須賀匠 *
Takumi Takasuka

高野雄紀 †
Yuki H. Takano

渡邊正太郎 ‡
Shotaro Watanabe

雲居玄道 *
Gendo Kumoi

Abstract— This study compares distance metrics for synoptic pattern classification using Self-Organizing Maps (SOM). We implemented Batch-SOM with eight metrics: Euclidean (EUC), Structural Similarity Index (5x5 window, SSIM5), Teweles-Wobus score (S1), and their fusion (S1+SSIM5). Using sea level pressure data from Japan (1991–2000), we evaluated performance via Macro-averaged Recall and medoid representation quality. Results showed S1+SSIM5 had superior generalization. We also introduced medoid and true medoid representations to reduce blurring from centroid averaging, enhancing the interpretability of SOM output maps for operational use, bridging advances in structural similarity metrics.

Keywords— Self-Organizing Map (SOM), synoptic pattern classification, structural similarity metrics

1 はじめに

総観規模の気圧配置パターンを分類する技術は、気象予報や気候システムの理解、さらには防災・減災といった多様な分野で基盤となるものである。従来の専門家による主観的な分類は、その知見を直接反映できる利点があるものの、多大な労力や再現性、拡張性に課題があった。そのため、客観的かつ自動的な分類手法が広く研究されている。

教師あり学習の分野では、木村ら [1] がサポートベクターマシン (SVM) を用いて「冬型」や「南高北低型」などを自動検出し、実用的な検索システムを構築した。しかし、このアプローチにはラベル付けのコストや、主観のばらつき (ノイズ) による学習データの品質の限界といった課題が指摘されている。

一方、教師なし学習では、自己組織化マップ (SOM) が総観パターンの持つ複雑な非線形構造を可視化・圧縮する有効な手段として用いられてきた [2, 3]。国内の研究では、筆者ら [4] が Batch-SOM を用いて日本周辺の気圧配置をクラスタリングし、1km メッシュの気象データと関連付ける研究成果を示している。

多くの従来研究ではユークリッド距離を基本としているが、勾配・形状・位置といった「構造」の評価には限界がある。これを改善するため、S-SOM[5] では構造類似

度指標 (SSIM) を BMU 探索に導入し、トポロジーの保存性を向上させた。佐藤らは、勾配ベースの S1 スコアと SSIM が人間の主観的な「類似性」の判断をよく再現することを統計的に実証した [6]。さらに、Winderlich らは改良 SSIM と medoid 表現を組み合わせ、クラス分離性と代表性を両立させる手法を提案している [7]。

本研究では、Batch-SOM フレームワーク上で 8 つの距離指標を厳密に比較する。既存手法として S1, SSIM5, Euclidean, 新たに提案する単一手法として KAPPA, GSSIM, そして複数の指標を統合した複手法を同一条件で評価する。従来の centroid 表現に代わり、medoid および true medoid を用いることで解釈性を向上させ、学習期 (1991–1997) と独立検証期 (1998–2000) を明確に分離して汎化性能を検証する。

2 従来手法

2.1 バッチ版 SOM

自己組織化マップ (SOM) [8] は教師なし学習によって高次元データを低次元 (通常 2 次元) マップに写像する手法である。オンライン版 SOM は 1 サンプルずつ処理するが、バッチ版 SOM (Batch-SOM) は全データまたはミニバッチ単位で処理することで計算効率と収束安定性を向上させる。

本研究では、Graphics Processing Unit (GPU) によるミニバッチ版バッチ SOM を実装した [9]。学習アルゴリズムは以下の通りである：

1. **初期化**： $M = m_x \times m_y$ 個のノード (ニューロン) を 2 次元格子に配置し、各ノードの重みベクトル $\mathbf{w}_j \in \mathbb{R}^D$ ($j = 1, \dots, M$) をランダムまたはデータからのサンプリングで初期化する。
2. **ミニバッチ処理**：各反復 t において、データをサイズ B のミニバッチに分割して処理する。
3. **BMU 探索**：各入力 \mathbf{x}_i に対して、最良一致ユニット (Best Matching Unit; BMU) を距離関数 $d(\cdot, \cdot)$ により決定：

$$c(i) = \arg \min_j d(\mathbf{x}_i, \mathbf{w}_j) \quad (1)$$

4. **近傍関数**：BMU $c(i)$ と各ノード j 間の近傍関数 $h_{c(i),j}$ を計算する。本研究ではガウス関数を使用：

$$h_{c(i),j} = \exp \left(-\frac{\|\mathbf{r}_{c(i)} - \mathbf{r}_j\|^2}{2\sigma(t)^2} \right) \quad (2)$$

* 〒 940-2188 新潟県長岡市上富岡町 1603-1 長岡技術科学大学, Nagaoka University of Technology, 1603-1 Kamitomiokamachi, Nagaoka, Niigata 940-2188, Japan. E-mail: s233319@stn.nagaokaut.ac.jp.

† 〒 305-0052 茨城県つくば市長峰 1-1 気象研究所, Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan.

‡ 〒 107-0052 東京都港区赤坂 5-4-7 The HEXAGON 5F 株式会社ウェザーマップ, Weather Map Co., Ltd., The HEXAGON 5F, 5-4-7 Akasaka, Minato-ku, Tokyo 107-0052, Japan.

ここで r_j はノード j のグリッド座標, $\sigma(t)$ は学習進行に応じて減衰する近傍幅である.

5. **重み更新**: ミニバッチ内の全サンプルについて分子・分母を累積し, 一括更新:

$$w_j^{new} = \frac{\sum_{i \in \text{batch}} h_{c(i),j} x_i}{\sum_{i \in \text{batch}} h_{c(i),j}} \quad (3)$$

6. **近傍幅の減衰**: $\sigma(t)$ は学習全体の進行に応じて単調減少させる. 本研究では漸近減衰を採用:

$$\sigma(t) = \frac{\sigma_0}{1 + t/(T/2)} \quad (4)$$

ここで σ_0 は初期近傍幅, T は総反復回数である.

この枠組みでは, 距離関数 $d(\cdot, \cdot)$ の選択が BMU 決定, ひいては学習結果全体に決定的な影響を与える. 以下では各距離関数を定義する.

2.2 距離関数

バッチ SOM および気象の気圧配置に関しては, 既存手法として以下の距離関数が用いられている. また, 本研究の距離関数に用いる記号を以下の通り定義する.

定義

- x, w : 入力パターンとノード重み ($H \times W$ の格子場)
- Ω : 格子領域の添字集合, $|\Omega|$ は格子点数
- $\mu_x(s), \mu_w(s)$: 画素 s 近傍 (5×5) での局所平均
- $\sigma_x^2(s), \sigma_w^2(s)$: 同局所分散
- $\text{cov}_{xw}(s)$: 同局所共分散
- Δ_x, Δ_y : 水平方向・鉛直方向の一次差分
- g_x, g_y : 内部共通格子での勾配成分, G : 勾配強度
- \mathbf{n} : 正規化勾配ベクトル, $\kappa(z) = \nabla \cdot \mathbf{n}$ (中心差分)
- ε : 数値安定化の微小正值 (10^{-12})

また, 以下の 3 種の距離関数は一部, 本研究に合わせて調整しているが本質的には佐藤ら [6] の定義した距離関数と同じ意味をもつ.

Euclidean (ユークリッド距離) 最も基本的な距離で, 値の差の総量を測る:

$$d_{\text{EUC}}(x, w) = \sqrt{\sum_{s \in \Omega} (x(s) - w(s))^2 + \varepsilon} \quad (5)$$

S1 (Teweles-Wobus Score) 一次差分 (勾配) の相対誤差に基づき, 前線帯の鋭さ・広がり・配置の一致を測る. 加法定数に不変:

$$d_{\text{S1}}(x, w) = 100 \times \frac{\sum_{(i,j)} |\Delta_x^x(i, j) - \Delta_x^w(i, j)| + \sum_{(i,j)} |\Delta_y^x(i, j) - \Delta_y^w(i, j)|}{\sum_{(i,j)} \max(|\Delta_x^x(i, j)|, |\Delta_x^w(i, j)|) + \sum_{(i,j)} \max(|\Delta_y^x(i, j)|, |\Delta_y^w(i, j)|) + \varepsilon} \quad (6)$$

ここで $\Delta_x^x(i, j) = x(i, j+1) - x(i, j)$, $\Delta_y^x(i, j) = x(i+1, j) - x(i, j)$, w 側も同様である.

SSIM5 (構造類似度指標) 5×5 の局所窓で平均・分散・共分散を計算し, 明るさ・コントラスト・構造の一致度を測る:

$$d_{\text{SSIM5}}(x, w) = 1 - \frac{1}{|\Omega|} \sum_{s \in \Omega} \frac{4 \mu_x(s) \mu_w(s) \sigma_x(s) \sigma_w(s) \text{cov}_{xw}(s)}{(\mu_x(s)^2 + \mu_w(s)^2)(\sigma_x^2(s) + \sigma_w^2(s))(\sigma_x(s) \sigma_w(s)) + \varepsilon} \quad (7)$$

$\mu, \sigma^2, \text{cov}$ は 5×5 移動窓内の局所統計量, 境界は reflect パディングである.

3 提案手法

本研究では, 各データ点は入力を各時刻の SLP 偏差 (hPa) をベクトル化したものとする. x, w はそれぞれ入力パターンとノード重みの 2 次元場 ($H \times W$), 格子領域を Ω とする. $\varepsilon = 10^{-12}$ は数値安定化のための微小正值である.

このデータ点に対して, 新たに以下の距離関数を提案する.

3.1 距離関数

GSSIM (勾配構造類似度) SSIM の概念を勾配場へ拡張する. 具体的には, (i) 勾配強度の一致 $\frac{2G^x G^w}{(G^x)^2 + (G^w)^2 + \varepsilon}$ と (ii) 方向の一致 $\frac{1}{2}(1 + \cos \theta)$ ($\cos \theta = \frac{g_x^x g_x^w + g_y^x g_y^w}{G^x G^w}$) を, $\max(G^x, G^w)$ で重み付けして強いエッジを重視する. 定式化は以下のとおりである.

$$d_{\text{GS}}(x, w) = 1 - \frac{\sum_{i,j} \max(G^x, G^w) \frac{2G^x G^w}{(G^x)^2 + (G^w)^2 + \varepsilon} \frac{1}{2} \left(1 + \frac{g_x^x g_x^w + g_y^x g_y^w}{G^x G^w + \varepsilon}\right)}{\sum_{i,j} \max(G^x, G^w) + \varepsilon} \quad (8)$$

g_x, g_y は内部格子の一次差分, $G = \sqrt{(g_x)^2 + (g_y)^2 + \varepsilon}$ である.

KAPPA (曲率距離) 正規化勾配の発散 (曲率) から等圧線の曲がり (凹凸・渦の閉じ具合) を比較した距離関数を以下に提案する.

$$d_{\text{KA}}(x, w) = \frac{1}{2} \times \frac{\sum_{i,j} |\kappa(x)(i, j) - \kappa(w)(i, j)|}{\sum_{i,j} \max(|\kappa(x)(i, j)|, |\kappa(w)(i, j)|) + \varepsilon} \quad (9)$$

$\kappa(z) = \nabla \cdot (\nabla z / (\|\nabla z\| + \varepsilon))$ は中心差分の内部共通格子で評価される.

3.2 複合手法

複合手法として, 異なる物理的意味を持つ距離指標を統合する新しい手法を提案する. 本研究の新規性は, 複数

の距離関数を Root Mean Square (RMS) で統合する点にある。

RMS 統合の理論的根拠 単純な算術平均や重み付き和ではなく RMS を採用した理由は以下の通りである：

1. **スケール不変性**：各距離指標は異なる物理量（勾配、曲率、構造類似度）を評価するため、そのスケールが大きく異なる。min-max 正規化により各指標を $[0, 1]$ 区間に標準化することで、スケールの違いを吸収する。
2. **ユークリッド的統合**：RMS は多次元空間における距離の自然な統合方法である。各正規化された距離を独立な次元と見なすと、RMS は多次元空間でのユークリッド距離に相当する：

$$d_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \tilde{d}_i^2} \quad (10)$$

3. **外れ値への感度**：算術平均と比較して、RMS は大きな値により敏感に反応する。これにより、いずれかの指標で大きな違いが検出された場合、統合距離もその違いを適切に反映する。
4. **係数フリー**：正規化により各指標が同じオーダー（ $[0, 1]$ 区間）になるため、重み係数を調整することなく等価的に統合できる。これは実用上の大きな利点である。

行方向正規化の意味 入力 x を固定し、全候補 w' に対して min-max 正規化することで、各入力に対する相対的な類似性を評価する。これにより、絶対的な距離値ではなく、候補間の相対的な順位関係に基づく BMU 選択が可能となる。

S1,KA S1（勾配の鋭さ）と KAPPA（曲率）を統合：

$$\tilde{d}_{\text{S1}}(x, w) = \frac{d_{\text{S1}}(x, w) - \min_{w'} d_{\text{S1}}(x, w')}{\max_{w'} d_{\text{S1}}(x, w') - \min_{w'} d_{\text{S1}}(x, w') + \varepsilon} \quad (11)$$

$$\tilde{d}_{\text{KA}}(x, w) = \frac{d_{\text{KA}}(x, w) - \min_{w'} d_{\text{KA}}(x, w')}{\max_{w'} d_{\text{KA}}(x, w') - \min_{w'} d_{\text{KA}}(x, w') + \varepsilon} \quad (12)$$

$$d_{\text{S1,KA}}(x, w) = \sqrt{\frac{\tilde{d}_{\text{S1}}(x, w)^2 + \tilde{d}_{\text{KA}}(x, w)^2}{2}} \quad (13)$$

GS,KA GSSIM（エッジの強さ・向き）と KAPPA（曲率）を統合：

$$d_{\text{GS,KA}}(x, w) = \sqrt{\frac{d_{\text{GS}}(x, w)^2 + \tilde{d}_{\text{KA}}(x, w)^2}{2}} \quad (14)$$

S1,GS,KA S1（鋭さ）・GSSIM（方向）・KAPPA（曲率）の三要素を総合：

$$d_{\text{S1,GS,KA}}(x, w) = \sqrt{\frac{\tilde{d}_{\text{S1}}(x, w)^2 + d_{\text{GS}}(x, w)^2 + \tilde{d}_{\text{KA}}(x, w)^2}{3}} \quad (15)$$

3.3 Medoid 表現

SOM の各ノードの代表パターンとして、従来は centroid（割り当てられたサンプルの平均）が使用されるが、平均化により勾配が平滑化される問題がある。そこで本研究では、各ノードに割り当てられたインデックス集合 I_c に対し、ノード内総距離

$$\text{cost}(i) = \sum_{j \in I_c} d(X_j, X_i) \quad (16)$$

が最小となる実サンプル X_{i^*} ($i^* = \arg \min_{i \in I_c} \text{cost}(i)$) をそのノードの true medoid とする。これにより実データの鋭さを保持した代表パターンが得られる。

4 実験設定

4.1 データと前処理

- **物理量**：海面更正気圧 (SLP)
- **領域**：日本周辺域 ($15\text{--}55^\circ\text{N}$, $115\text{--}155^\circ\text{E}$)
- **期間**：学習 1991–1997 年、検証 1998–2000 年（日次 09 Japan Standard Time; JST）
- **ラベル**：吉野 [10] の「気圧配置ごよみ」準拠
- **前処理**：Pa から hPa へ変換後、領域平均を引く。

4.2 ラベル体系の詳細

本研究で用いるラベル体系は、吉野 [10] の付録 B「気圧配置ごよみ」に準拠する。基本型は以下の 6 分類（計 15 サブタイプ）で構成される：

1. **西高東低冬型** 冬季に特徴的なシベリア高気圧とアリューシャン低気圧に起因する西高東低の場。
2. **気圧の谷型** 低気圧あるいは気圧の谷の通過に対応する型。
 - A. 低気圧が北海道またはサハリン付近を東に進む
 - B. 低気圧が日本海から北東に進む
 - C. 低気圧が台湾から日本の太平洋岸を東～東北東に進む
 - D. 二つ玉低気圧、または日本海と太平洋に低圧部
3. **移動性高気圧型** 移動性高気圧に覆われる型。
 - A. 日本の北方または北部を東に進む
 - B. 日本列島上（主として本州上）を東に進む
 - C. 帯状高気圧
 - D. 日本の太平洋岸または南方を東に進む
4. **前線型** 停滞性の前線が卓越する型。
 - A. 日本列島上をほぼ東西方向に走る主として停滞性の前線
 - B. 太平洋岸または日本南方をほぼ東西方向に走る主として停滞性の前線
5. **南高北低夏型** 原則として北太平洋高気圧が日本列島を支配する夏型。
6. **台風型** 台風の位置に応じて次のサブタイプに区分する。

- A. 台風が南九州より南方の海上にある場合
- B. 台風が本州およびその沿岸地帯にある場合
- C. 台風が北日本にある場合

この他に、移行型（例：「3A - 2D」）と複合型（例：「2A + 2C」）が定義されているが、本研究では基本型 15 種のみを評価対象とした。詳細な定義は吉野 [10] を参照されたい。

4.3 SOM 設定と学習条件

- マップサイズ：10 × 10
- 反復回数：1000, バッチサイズ：128
- 近傍関数：ガウス関数, 初期 $\sigma = 2.5$ （学習全体で減衰）
- 学習率：0.5

4.4 評価指標

- Macro Recall（基本ラベル）：各基本ラベルの再現率の平均。混同行列は「基本ラベル vs クラスタ列」で集計する。
- NodewiseMatchRate（基本）：各ノードの多数決（基本ラベル）と true medoid の基本ラベルの一致率。

学習期および独立検証期に対して算出する。

5 実験結果

本節では、シード値 1~50 で 50 回試行した集計結果を示す。手法別の Macro Recall（学習・検証）および NodewiseMatchRate（学習）の要約統計を表 1 に、基本ラベルごとの平均再現率の横断ピボット表を学習（表 2）と検証（表 3）でそれぞれ示す。

50 試行において学習期の Macro Recall は S1,KA / S1,GS,KA / GS,KA が 0.37 ± 0.02 で同等の上位値を示した。検証期では S1,KA が 0.23 ± 0.02 でやや高く、実用上は S1,KA が汎化性能において優位と判断できる。NodewiseMatchRate（学習）は複合手法のスコアが低い。

加えて、各手法が「どのくらいのラベルを検出できているか（非ゼロ再現率の基本ラベル数）」に着目すると、学習期では SSIM5 のみが 15/15 ラベルを網羅し、他の 7 手法はいずれも 14/15（主として 6C が 0）にとどまった（表 2）。一方で検証期は差がより明瞭で、S1 が 11/15 と最も広いカバレッジを示し、S1,GS,KA・GS,KA・GSSIM が 10/15, S1,KA が 8/15, KAPPA / Euclidean / SSIM5 はいずれも 7/15 にとどまった（表 3）。

検出が安定なラベルとしては、1（西高東低冬型）、3B（移動性高気圧：本州上）、4A（東西前線）が挙げられ、多くの手法で検証期でも高い再現率が維持される。一方で、2C（太平洋岸低気圧）、3C（帯状高気圧）、3D（南岸高気圧）、台風系（6A/6C）は複数手法で検出ゼロになりやすく、事例数の少なさや場の多様性の高さが影響していると解釈できる。

以上より、実運用では S1,KA を標準距離とする構成が妥当と考える。

図 2 に、S1,KA 手法における学習精度が最も高かった seed=20 の SOM 出力（10 × 10）を示す。左は centroid（平均）であり、平均化の影響により低気圧中心や前線の鋭さが弱まり、いわゆる「雪だるま現象」によりノードの気圧配置に対するラベル付けが困難となる。特に台風（6 系）では極小が鈍化し識別がほぼ不可能である。一方、右の true medoid では実データの鋭さ・位相が保持され、ノード多数決ラベルと medoid の基本ラベルが一致したノードを色付け（1= 青, 2D= 黄色, 3B= オレンジ, 4A= ピンク, 4B= 紫, 5= 赤, 6B= 緑）することで、専門家の手を借りずとも各ノードに対して一貫したラベル付けが可能となることが期待される。

6 考察

6.1 手法の特性と補完性

S1 は勾配差の比率に基づき前線帯や冬型で有効である。特に等圧線の密集度や前線の鋭さを定量化する点で気象学的に妥当性が高い。SSIM5 は局所構造（形状・位置・コントラスト）を評価し、人間の視覚認知と整合する。GSSIM は勾配強度と方向の両面を重み付きで評価し、エッジ検出の観点から構造を捉える。KAPPA は等圧面の曲率構造に注目し、高低気圧の閉じた構造や渦度場の特徴を抽出する。

複合手法（S1,KA / GS,KA / S1,GS,KA）はこれらの利点を RMS 統合で併合し、BMU 探索における汎化を支える。特に S1,KA は勾配（前線の鋭さ）と曲率（高低気圧の構造）の両面を捉えることで、総観規模現象の本質的特徴を包括的に評価できる。

6.2 Medoid 表現の有効性

centroid は平均化により勾配が平滑化されやすい一方、medoid は実データ実現値を代表として保持し、解釈性と検索性を高める。本実装では標準で true medoid を算出する。

6.3 汎化性能の解釈

S1 は勾配差に敏感、KAPPA は曲率（幾何）を評価するため、S1,KA は形状と鋭さを同時に捉え BMU 探索の頑健性を高める。その結果、検証 Macro Recall で最良を示した。一方、NodewiseMatchRate では Euclidean が最大であり、ノード内多数決と medoid ラベルの一致を生みやすい（振幅差に整合）。用途により、識別（Macro Recall）優先か、ノード代表性（一致率）優先かで距離の使い分けが必要である。

6.4 ラベル不均衡と評価設計

台風系（6A/6B/6C）は出現頻度が極端に低く、さらに複合場が約半数を占めるため、検証で再現率が低下しやすい。季節別 SOM（冬・梅雨・台風季）や階層的分類（大

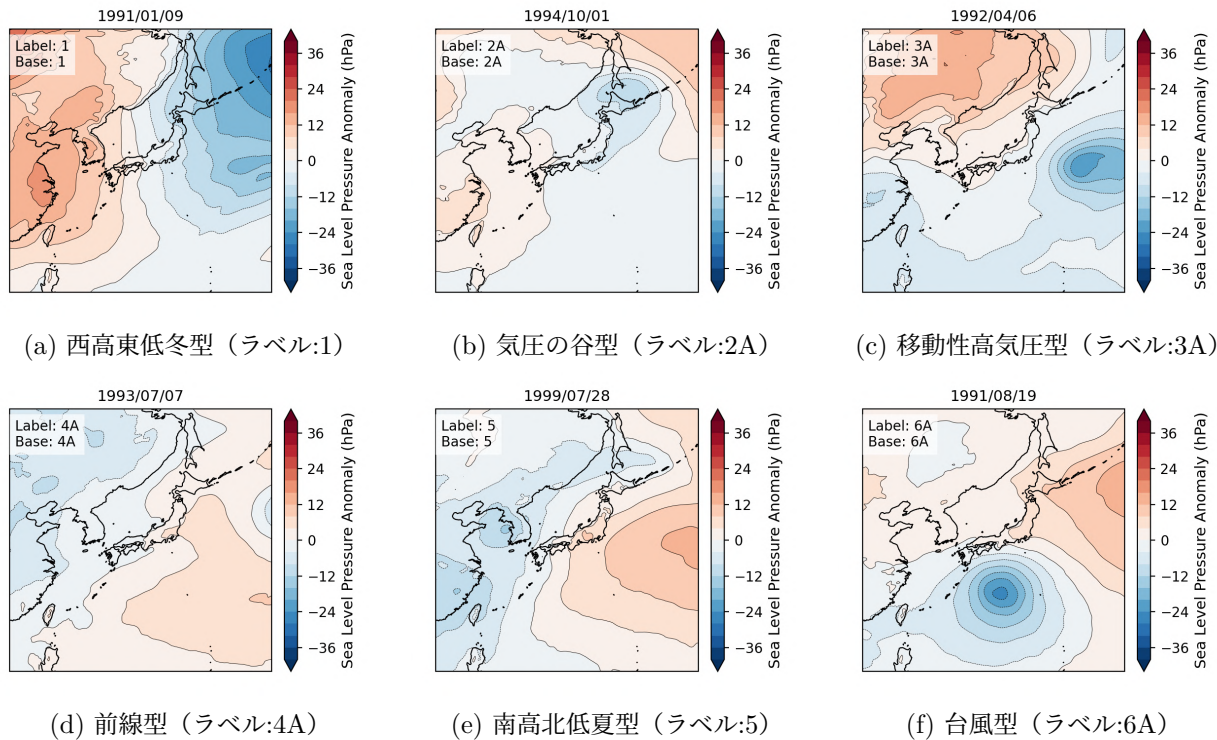


図1 基本型の代表的な海面更正気圧偏差パターン (単位: hPa). 各パネルは日本周辺域 (15–55°N, 115–155°E) における領域平均からの偏差を示す。

表1 手法別 Macro Recall (学習・検証) と NodewiseMatchRate (学習) の統計量 (平均 ± 標準偏差, N=50)

Method	Macro Recall (学習)	Macro Recall (検証)	NodewiseMatchRate (学習)
S1,KA	0.37 ± 0.02	0.23 ± 0.02	0.25 ± 0.03
S1,GS,KA	0.37 ± 0.02	0.22 ± 0.02	0.24 ± 0.04
GS,KA	0.37 ± 0.02	0.21 ± 0.01	0.25 ± 0.05
S1	0.36 ± 0.02	0.21 ± 0.02	0.26 ± 0.04
GSSIM	0.35 ± 0.02	0.21 ± 0.02	0.27 ± 0.04
SSIM5	0.32 ± 0.02	0.17 ± 0.01	0.29 ± 0.04
Euclidean	0.30 ± 0.02	0.18 ± 0.01	0.31 ± 0.03
KAPPA	0.30 ± 0.01	0.21 ± 0.02	0.29 ± 0.03

表2 横断ピボット表 (学習: 基本ラベルごとの平均再現率)

Method	1	2A	2B	2C	2D	3A	3B	3C	3D	4A	4B	5	6A	6B	6C	Overall	N	lab
S1,KA	0.92	0.26	0.24	0.14	0.60	0.21	0.79	0.05	0.01	0.68	0.35	0.59	0.61	0.10	0.00	0.37	14	
S1,GS,KA	0.91	0.35	0.27	0.07	0.61	0.20	0.79	0.08	0.02	0.66	0.34	0.59	0.45	0.17	0.00	0.37	14	
GS,KA	0.90	0.32	0.27	0.07	0.62	0.21	0.79	0.07	0.02	0.66	0.35	0.63	0.41	0.17	0.00	0.37	14	
S1	0.91	0.29	0.21	0.13	0.61	0.22	0.77	0.05	0.02	0.66	0.37	0.54	0.49	0.20	0.00	0.36	14	
GSSIM	0.91	0.27	0.18	0.09	0.59	0.19	0.75	0.06	0.02	0.63	0.31	0.55	0.56	0.18	0.00	0.35	14	
SSIM5	0.90	0.13	0.07	0.17	0.45	0.22	0.70	0.06	0.16	0.59	0.36	0.44	0.42	0.14	0.04	0.32	15	
Euclidean	0.91	0.22	0.12	0.05	0.48	0.35	0.74	0.05	0.14	0.57	0.22	0.33	0.34	0.06	0.00	0.30	14	
KAPPA	0.88	0.25	0.22	0.10	0.45	0.06	0.78	0.07	0.01	0.68	0.26	0.62	0.04	0.04	0.00	0.30	14	

表3 横断ピボット表 (検証: 基本ラベルごとの平均再現率)

Method	1	2A	2B	2C	2D	3A	3B	3C	3D	4A	4B	5	6A	6B	6C	Overall	N	lab
S1,KA	0.91	0.21	0.03	0.00	0.57	0.00	0.80	0.00	0.00	0.73	0.08	0.09	0.00	0.00	0.00	0.23	8	
S1,GS,KA	0.92	0.20	0.02	0.00	0.45	0.02	0.80	0.00	0.00	0.64	0.08	0.12	0.00	0.01	0.00	0.22	10	
GSSIM	0.93	0.09	0.02	0.00	0.48	0.02	0.78	0.00	0.00	0.53	0.10	0.14	0.00	0.11	0.00	0.21	10	
GS,KA	0.93	0.14	0.04	0.00	0.44	0.03	0.79	0.00	0.00	0.62	0.08	0.14	0.00	0.00	0.00	0.21	10	
S1	0.95	0.04	0.01	0.00	0.53	0.01	0.77	0.00	0.00	0.65	0.06	0.08	0.02	0.04	0.00	0.21	11	
KAPPA	0.91	0.15	0.00	0.00	0.31	0.00	0.73	0.00	0.00	0.74	0.05	0.20	0.00	0.00	0.00	0.21	7	
Euclidean	0.86	0.04	0.00	0.00	0.29	0.00	0.83	0.00	0.00	0.66	0.01	0.01	0.00	0.00	0.00	0.18	7	
SSIM5	0.86	0.00	0.00	0.01	0.28	0.00	0.78	0.00	0.00	0.52	0.08	0.05	0.00	0.00	0.00	0.17	7	

局型 → サブタイプ), 外部事象フラグの併用 (台風客観解析) により稀少クラスの頑健性を補うことができる。

7 まとめと今後の課題

Batch-SOM 上で, 既存手法 (S1, SSIM5, Euclidean) と提案手法 (KAPPA, GSSIM, S1, KA, GS, KA, S1, GS,

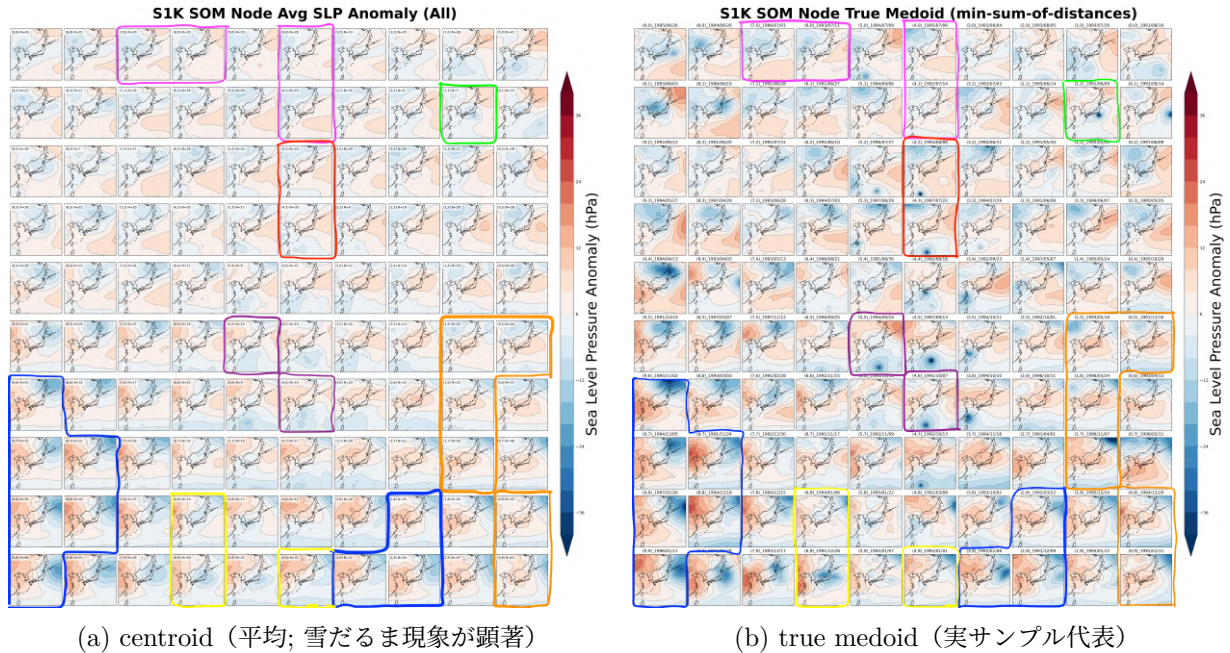


図2 S1,KA 手法における最良学習 (seed=20) の 10 × 10 SOM 出力マップ。右図 (true medoid) では、ノード多数決ラベルと medoid の基本ラベルが一致したノードに色付けしている (凡例: 1= 青, 2D= 黄色, 3B= オレンジ, 4A= ピンク, 4B= 紫, 5= 赤, 6B= 緑)。

KA) の 8 手法を同一条件で比較できる実験枠組みを構築した。実装に準拠した厳密な数式定義を提示し、評価は基本ラベルに対する Macro Recall とノード代表一致率で実施した。centroid の代わりに medoid を採用することで代表パターンの鋭さを保持できる。S1,KA のスコアが高く実務では S1,KA の運用が望ましい。今後の課題は、多変量拡張、季節別最適化などである。

参考文献

- [1] 木村広希, 川島英之, 日下博幸, 北川博之, “サポートベクターマシンを用いた気圧配置検出手法の提案—西高東低冬型を対象として—,” 地理学評論 Series A, 2009.
- [2] K. Philippopoulos, D. Deligiorgi, and G. Kouroupetroglou, “Performance comparison of self-organizing maps and k-means clustering techniques for atmospheric circulation classification,” methods, 2014.
- [3] N. Jiang, K.N. Dirks, and K. Luo, “Classification of synoptic weather types using the self-organising map and its application to climate and air quality data visualisation,” Weather and Climate, 2013.
- [4] 高須賀匠, 高野雄紀, 渡邊正太郎, 雲居玄道, “自己組織化マップを用いた気圧配置のクラスタリングと 1km メッシュ天気データによる分析,” 情報処理学会全国大会, 2024.
- [5] Q.-V. Doan, H. Kusaka, T. Sato, and F. Chen,

“S-som v1. 0: a structural self-organizing map algorithm for weather typing,”

Geoscientific Model Development, 2021.

- [6] S. Takuto and K. Hiroyuki, “Statistical intercomparison of similarity metrics in sea level pressure pattern classification,” 気象集誌. 第 2 輯, 2021.
- [7] K. Winderlich, C. Dalelane, and A. Walter, “Classification of synoptic circulation patterns with a two-stage clustering algorithm using the modified structural similarity index metric (ssim),” Earth System Dynamics, 2024.
- [8] T. Kohonen, “The self-organizing map,” Proceedings of the IEEE, 1990.
- [9] G. Vettigli, “Minisom: minimalistic and numpy-based implementation of the self organizing map,” 2018. <https://github.com/JustGlowing/minisom/>
- [10] 正敏吉野, 日本の気候—最新データでメカニズムを考える, 二宮書店, 2002.