

日本気象学会

九州支部だより

No. 147 2024年2月



発行者
日本気象学会九州支部
〒810-0052
福岡市中央区大濠1-2-36
福岡管区気象台
地域防災推進課内

Tel: 092-725-3614
Mail: info@msj-kyushu.jp
HP: <http://msj-kyushu.jp/>

今回の記事

◆サイエンスカフェ(2/3)開催報告

◆支部会員だより

「帰納的機械学習モデルを用いた アリソフの全球気候区分図の改訂」

熊本大学大学院 自然科学教育部 島袋 琉

熊本大学大学院 先端科学研究部 富田 智彦

サイエンスカフェ(2/3)開催報告

九州支部事務局

「地球温暖化と農業」をテーマに、令和6年2月3日(土)に第14回のサイエンスカフェを、4年ぶりに対面で開催し、14名の方が参加しました。専門家の広田先生の知識とユーモアに溢れた話し方や、案内人の平河さんの絶妙な進行で、とてもいい雰囲気ではが進んでいきました。会場から多数質問もあり、大盛況で終わることが出来ました。



左:平河さん(気象予報士会西部支部) 右:広田教授(九州大学大学院農学研究院)

支部会員だより

帰納的機械学習モデルを用いたアリソフの全球気候区分図の改訂

熊本大学大学院 自然科学教育部 島袋 琉
熊本大学大学院 先端科学研究部 富田 智彦

気候区分の考え方は、気候をよく反映していると思われる植生・土壌の分布に基づいて分類される経験的気候区分と、気候を形成する大気大循環・気団の分布に基づいて分類される成因的気候区分とに大別されます(水越・山下 1985)。本研究は、代表的な成因的気候区分の1つである、アリソフの気候区分図の改訂について簡単に紹介します。詳細については、下記の原著論文をご一読ください。

Shimabukuro, R., T. Tomita and K. Fukui, 2023: Update of global maps of Alisov's climate classification. *Progress in Earth and Planetary Science*, **10**, 1-18. <https://doi.org/10.1186/s40645-023-00547-1>

1. はじめに

一般に気候は、気温と水蒸気量が水平方向にほぼ一様な大規模の空気塊(気団)によって決定されます。地球規模の気団帯と前線の平均的な位置は、大気大循環の季節変化と対応し、季節的に南北に移動します。アリソフ(1954)は、こうした気団帯や前線の季節的な南北変動に着目し、1年中同じ気団帯に支配される4つの安定型気候帯と、季節により気団帯の交代が生じる3つの季節推移型気候帯の分類を考えました。まずアリソフは、世界の大気を地理的緯度すなわち気温によって赤道気団帯、熱帯気団帯、寒帯気団帯、極気団帯の4つに分け、次に各気団帯の1月(北半球の冬)と7月(北半球の夏)の分布域を重ね合わせることで、各季節の気団帯分布に基づく7つの気候帯を設定しています(図1)。アリソフ(1954)によれば、サハラ砂漠などは1年中、高温・乾燥の大気特性を持つ熱帯気団帯に覆われた熱帯気候帯に含まれ(ただし、ここでいう熱帯気団は現在の亜熱帯高気圧、すなわち亜熱帯気団に該当する)、この緯度帯では前線が降雨に影響を与えないために砂漠化をもたらすという成因を説明しています(Alisov 1954)。

しかし実際には、アリソフの気候区分図(図1)はかなり模式的な図であり、夏に降水量の少ないローマと降水量の多い東京が同じ気候帯に含まれているなど、実際の気候分布とは必ずしもよく対応しないという欠点が指摘されています(仁科 2019)。

また鈴木(1961)は、主前線の位置が正しく決定されているのかどうかについて疑問を呈しています。さらにアリソフ(1954)は、地表面の状態により大陸性／海洋性、大気大循環の特性により東岸型／西岸型、地形により平地型／山地型の類型によって、7つの気候帯から22の気候区へと細分化することを試みていますが、その際の具体的な手続きは示されておらず、また図示化も行われていません(水越・山下 1985)。このように、1950年代になされたアリソフを始めとする先人たちの気団論に基づく気候分類では、客観的な気団域の判定と気候帯の細分化に大きな課題が残っていました。本研究では、全球の再解析データに機械学習のクラスタリング技術を適用し、気団域を定量的かつ客観的に決定し、現在の気候をより現実的に表現できる、新しいアリソフの気候区分図へと約70年ぶりに更新しました。

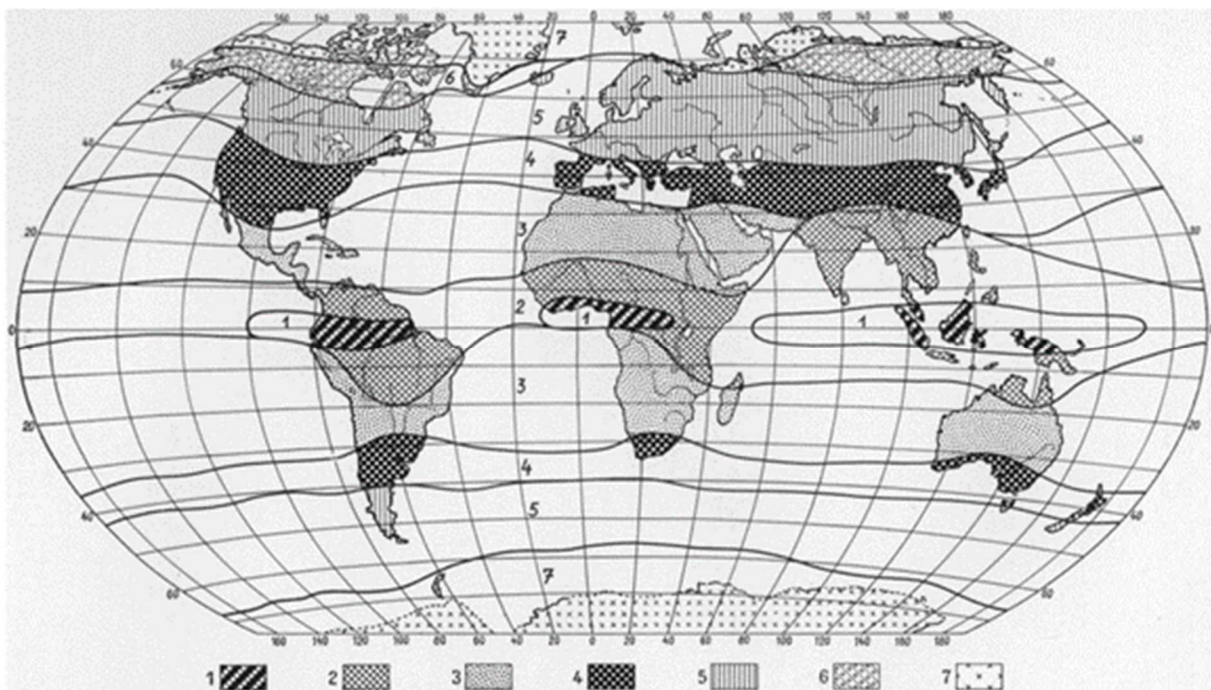


図1 アリソフの7気候帯(Alisov 1954)。図はFlohn 1957より引用。1:赤道気候帯、2:亜赤道気候帯、3:熱帯気候帯、4:亜熱帯気候帯、5:寒帯気候帯、6:亜寒帯気候帯、7:北極／南極気候帯。

2. データと解析方法

本研究では、次に挙げる2つのデータを使用しています。(1) ERA5の再解析データ(Hersbach et al. 2020)。物理量は対流圏下層高度(925–775 hPa)の気温(T ; 単位K)と比湿(q ; 単位 g kg^{-1})。領域は全球で空間解像度は 0.25° 。期間は1981–2020年(40年間)の月平均データ。 T と q は、気団を特徴づけるための気候変数で、クラスタリングの特徴量として用います。(2) GPCP Ver. 2.3(Adler et al. 2018)の降水量データ(単位 mm day^{-1})。領域は全球で空間解像度は 2.5° 。期間ならびに時間解像度はERA5と同様です。降水量は、 T と q のクラスタリングによって決定された気団帯分布の境界、すなわち気候学的前線の位置を診断するために用います。

水平方向の大気特性の拡がりに加え、鉛直方向にも均質な拡がりを持つ3次元的な大規模空気塊として気団を捉えるため、気候変数 x の 925–775 hPaにおける鉛直平均 $\bar{x}$ を次式で定義します:

$$\bar{x} \equiv \frac{1}{g} \int_{775\text{hPa}}^{925\text{hPa}} x \, dp \bigg/ \frac{1}{g} \int_{775\text{hPa}}^{925\text{hPa}} dp$$

ここで、 g は重力加速度、 p は気圧、 x は T (K) または q (g kg^{-1}) を表しており、上式の分子は鉛直方向の空気密度変化を考慮した、 x の質量重み付き鉛直積算です。変動性の高い小規模擾乱の影響を取り除くために、10–3月(北半球の寒候季)と4–9月(北半球の暖候季)で両気候変数の6ヶ月平均値を取ります(図2)。次に両変数の単位尺度を揃えるため、2季節のデータを合わせた全球の水平格子点値の平均と標準偏差を計算し、これらの値で特徴量を標準化(平均0、標準偏差1に変換)しました。

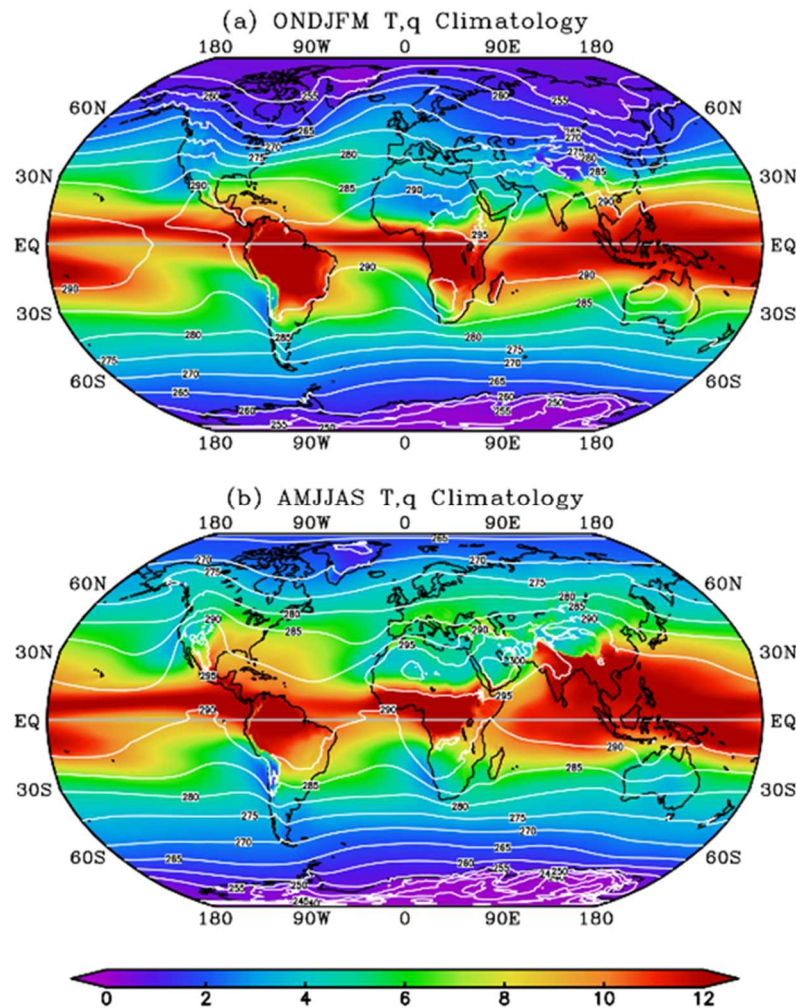


図2 1981–2020年(40年間)の925–775hPaで鉛直平均した(a) 10–3月(北半球の寒候季)、(b) 4–9月(北半球の暖候季)の気温(単位K、等値線、等値線間隔は5 K)と比湿(単位 g kg^{-1} 、色、色のスケールは図下の通り、暖色系・寒色系はそれぞれ相対的に高い・低い水蒸気量を示す)の気候値の分布。

クラスタ(=気団帯)の季節による基準を揃えるために、両半年全体で標準化したデータを入力データとして同時に与えました。従って、再解析データの特徴量の数、格子点の数、季節の数を掛け合わせると、入力データの配列サイズは $2(T, q) \times 1440(\text{経度数}) \times 721(\text{緯度数}) \times 2(\text{寒候季、暖候季}) = 4,152,960$ となります。次に、このデータに対してK-means++法(Arthur and Vassilvitskii 2007)を適用しました。K-means法(MacQueen 1967)は、代表的な非階層的クラスタリング手法であり、産業界や学術界において広く利用されています。K-means法は、近いクラスタを順次統合していくWard法のような階層型手法に比べて距離計算の量が少なく済むため、ビッグデータの解析に適しています。さらにK-means++法は、分類結果が初期の重心位置に依存する問題があった従来のK-means法に比べ、分類がより効果的になされ分類に一貫性のあることが報告されています(Raschka and Mirjalili 2020)。従って本研究でも、K-means法やK-medoids法のような非階層型手法の初期値依存性問題を克服した、K-means++法を採用しました。初期値の依存性をさらに小さくするため、初期の重心位置を変えてK-means++法を20回実行し、クラスタ内の誤差平方和(歪み)が最も小さくなったモデルを最適なクラスタリングモデルとして最終的に選択しました。

階層型や密度ベースのクラスタリング手法ではなく、K-means++法を用いる利点の一つとして、特徴空間における重心の最終的な座標位置を保持しておき、異なるデータセットに対しても同じ基準で分類できる点があります。この利点は、地球温暖化に伴う気団域・気候学的前線の位置変化を追跡することを可能にします。しかし、K-means++法における最大の問題は、最適なクラスタ数を事前に設定する必要があることです。本研究では、4つの定量的な評価指標を用いて、最適なクラスタ数(ここでは気団帯の数に相当)を決定します。

1つ目の指標は、Davies-Bouldin 指数(DBI; Davies and Bouldin 1979)であり、DBIは以下の式で定義されます:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \left(\frac{s_i + s_j}{d_{i,j}} \right)$$

ここで、 K はクラスタ数、 s_i はクラスタ c_i に属する全サンプル間の平均距離、 $d_{i,j} = \|c_i - c_j\|$ は2つのクラスタ c_i と c_j の重心間のL2ノルム(ユークリッド距離)です。DBIは、 $i, j = 1, \dots, K$ のクラスタ c_i とそれに最も近いクラスタ c_j との間の平均類似度を表します。DBIの値が小さいほど、最近接のクラスタのペア c_i と c_j を隔てる距離が大きいことを示し、最適なクラスタ数であることを示します。

2つ目の指標は、シルエット係数(SC;Rousseeuw 1987)であり、単一サンプル a_n に対するシルエット係数 SC_n は、以下の式で定義されます:

$$SC_n = \frac{e_n - f_n}{\max(e_n, f_n)}$$

ここで、 f_n は a_n と同じクラスタ内に属する全サンプル間の平均距離、 e_n は a_n が属するクラスタと最も近い別のクラスタに属する全サンプル間の平均距離を表します。 SC_n の値は $-1 \sim 1$ の範囲で変化し、1に近いほど分類の性能が良いことを示します。全サンプルデータについてのシルエット係数 $SC = \sum_{n=1}^N SC_n / N$ (N は全サンプル数)はクラスタリングモデル全体の性能を説明するために用いられます。

3つ目の指標は、Caliński-Harabasz 指数(CHI;Caliński and Harabasz 1974)であり、CHIは以下の式で定義されます:

$$CHI = \frac{\text{trace}(B_K)}{\text{trace}(W_K)} \times \frac{N - K}{K - 1}$$

ここで、 $B_K = \sum_{i=1}^K n_i (c_i - c_o)(c_i - c_o)^t$ はクラスタ間の分散行列(n_i はクラスタ C_i に属するサンプル数、 c_o は全てのサンプルデータの重心)、 $W_K = \sum_{i=1}^K \sum_{a_n \in C_i} (a_n - c_i)(a_n - c_i)^t$ はクラスタ内の分散行列です。上付き文字 t は転置を表します。CHIの値が大きいほど、各クラスタに属するサンプル分布は密であり、クラスタ間の距離が大きいことを示し、最適なクラスタ数であることを示します。

4つ目の指標は、有限のモデル集合からの統計的なモデル選択に用いられる、ベイズ情報量規準(BIC;Schwarz 1978)です。クラスタ C_i のサンプルがガウス分布に従うという仮定の下、BICは以下のように定義されます:

$$BIC = \sum_{i=1}^K \left[\log L(\hat{\theta}_i; a_n \in C_i) - \frac{D}{2} \log n_i \right]$$

ここで、 L はモデルの尤度関数、 $\hat{\theta}_i$ は尤度関数のパラメータセット、 D は K 個の成分を持つモデルのパラメータ数です。第2項は、パラメータ数の増加によるモデルの過学習を防ぐために課されるペナルティ項を表します。ただし、ここではBICを通常の設定とは逆の符号で計算していることに注意して下さい。よって、BICの値が大きいほど、サンプルデータや未知データの分布に対するモデルの適合度が高いことを表し、最適なクラスタ数であることを示します。

3.1. 最適なクラスタ数(気団帯の数)の決定

図3は、クラスタ数(＃と略す)に対する評価指標の値を示しています。DBI(青)は、＃2～＃4までクラスタ数が増えるにつれて値が大きく減少し、＃4で最小値を取った後、＃5以降で増加しています。このことは、＃4が最適なクラスタ数であることを示します。SC(赤)は、＃2～＃6までクラスタ数が増えるにつれて値が大きく減少し、＃2で最も良い結果を示していますが、＃3～＃4ではクラスタリングの分類性能をほぼ維持しています。CHI(緑)とBIC(橙)は、＃2～＃10までクラスタ数が増加するにつれ、両者ともに増加する傾向にありますが、＃4を境にして変化の割合が減少しています。これらの指標を総合的に判断し、最適なクラスタ数は＃4であると判断します。このことは、地球の気候は統計的には4つの気団帯に分けられることを示しています。つまり、アリソフによって1950年代になされた4つの分類は、現在の高品質な再解析データを用いた、データ駆動型の観点からも支持される結果となりました。

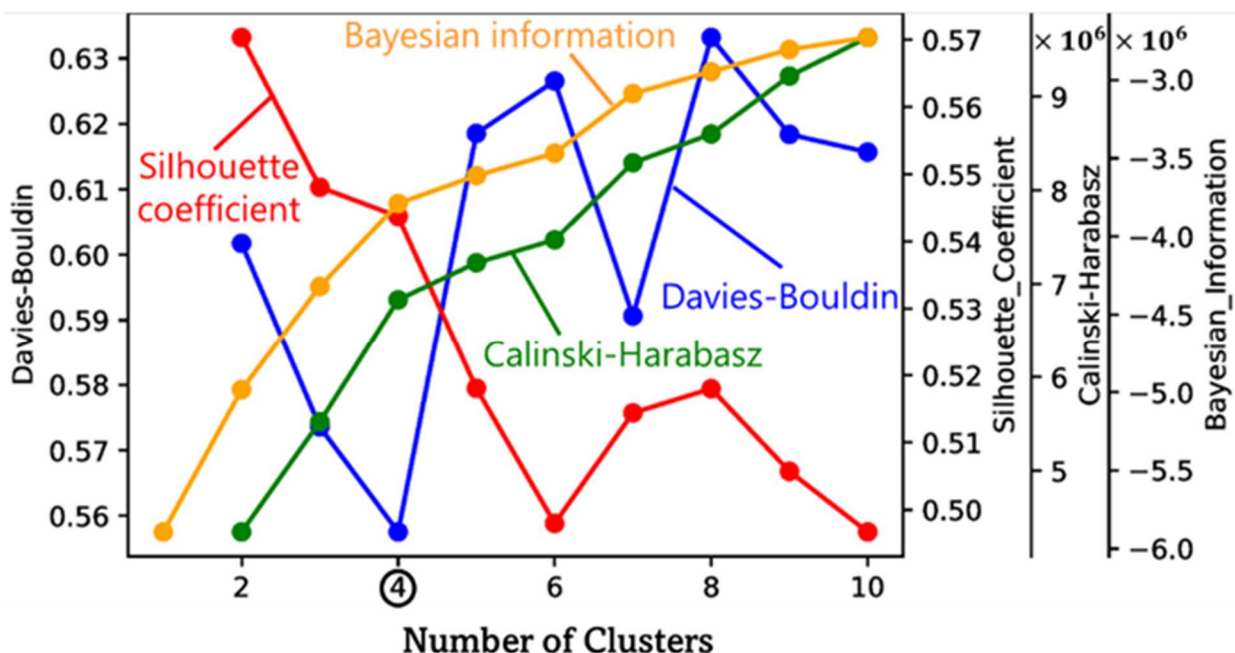


図3 最適なクラスタ数を決定する4つの定量的評価指標：DBI(青)、SC(赤)、CHI(緑)、BIC(オレンジ)。各指標の説明は本文参照。指標は全て無次元。

3.2. 気団帯の全球分布域と9気候帯の設定

図4はK-means++法を用いて、世界の気候を4つのクラスタ(気団帯)に分類した結果を示しています。クラスタリングによって決定された各気団帯を、低緯度側から高緯度側にかけて、熱帯気団帯(赤色)、亜熱帯気団帯(黄色)、寒帯気団帯(水色)、極気団帯(紫色)とそれぞれ名付けます。ただし、ここでいう熱帯・亜熱帯気団帯は、アリソフ(1954)でいう赤道・熱帯気団帯にそれぞれ対応していることに注意します。季節変化に伴って

各気団帯は夏半球では極側へ（図 4b）、冬半球では赤道側へと移動し（図 4a）、地球規模の気団帯および前線の季節的な南北移動を捉えています。また、クラスタリングによる客観的な気団判別によると、暖候季の北極圏では極気団帯は出現せず、寒帯気団帯が最北域に広がっています（図 4b）。

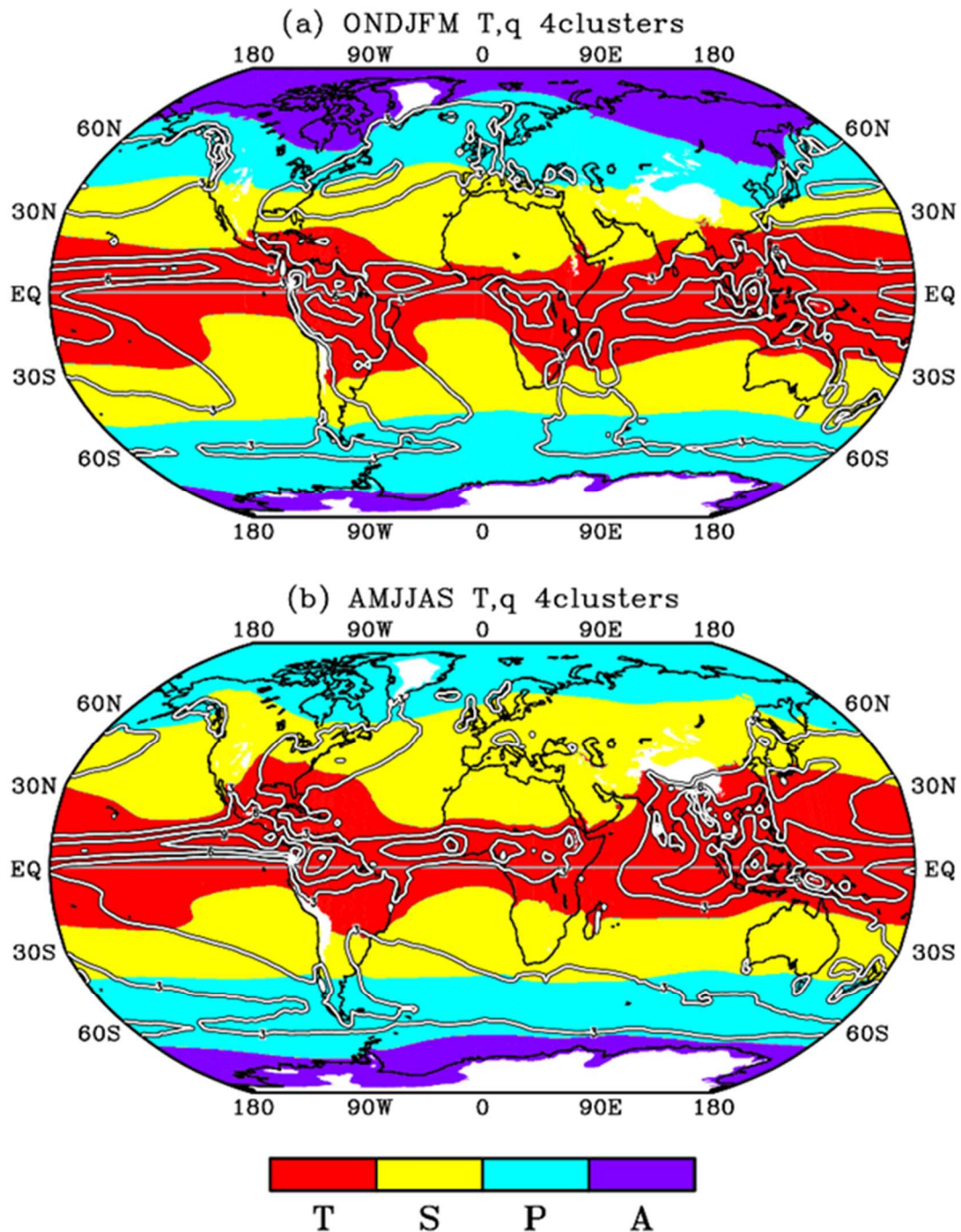


図4 (a) 10-3月、(b) 4-9月における4つの気団帯（色）と 3 mm day^{-1} 以上の強い降水量（白い等値線、等値線の間隔は 3 mm day^{-1} ）の全球分布。空白域は高度約2,000m以上（775 hPa以下）の領域を表す。T：熱帯（Tropical）、S：亜熱帯（Subtropical）、P：寒帯（Polar）、A：極（Arctic / Antarctic）。

次に、気候学的前線の位置を強い降水域の位置と比較して診断します。図4の白い等値線は、降水量が 3 mm day^{-1} 以上の強い降水域を示します。先ず、世界の強い降水域のほとんどは熱帯気団帯の中に現れていることが分かります。それ以外の中・高緯度における降水域に注目してみると、北半球の冬季では、北大西洋上と日本の東方海上において、亜熱帯－寒帯気団帯の境界（寒帯前線）と海上の強い降水域がうまく一致していることが分かります（図4a）。このことは、機械学習のクラスタリング技術が、傾圧不安定に伴う温帯低気圧の発生・発達経路または前線性の降水をうまく捉えることができることを表し、気団気候学においてこの手法が有効であることを示します。また北半球の暖候季では、北半球における熱帯－亜熱帯気団帯の境界（亜熱帯前線）と日本付近の強い降水バンドが一致しており（図4b）、相当温位の南北勾配で特徴づけられる梅雨前線と対応すると考えられます。ここでのクラスタリングは、従来のアリソフの気候区分図（図1）で指摘されていた、中・高緯度の強い前線性の降水域をより正確に捉えています。

図5はアリソフ（1954）の手法に倣って、寒候季と暖候季における気団帯の分布域（図4a, b）を重ね合わせて作成した9気候帯の図であり、表1はその気候帯の特徴をまとめています。従来のアリソフ気候区分図（図1）と

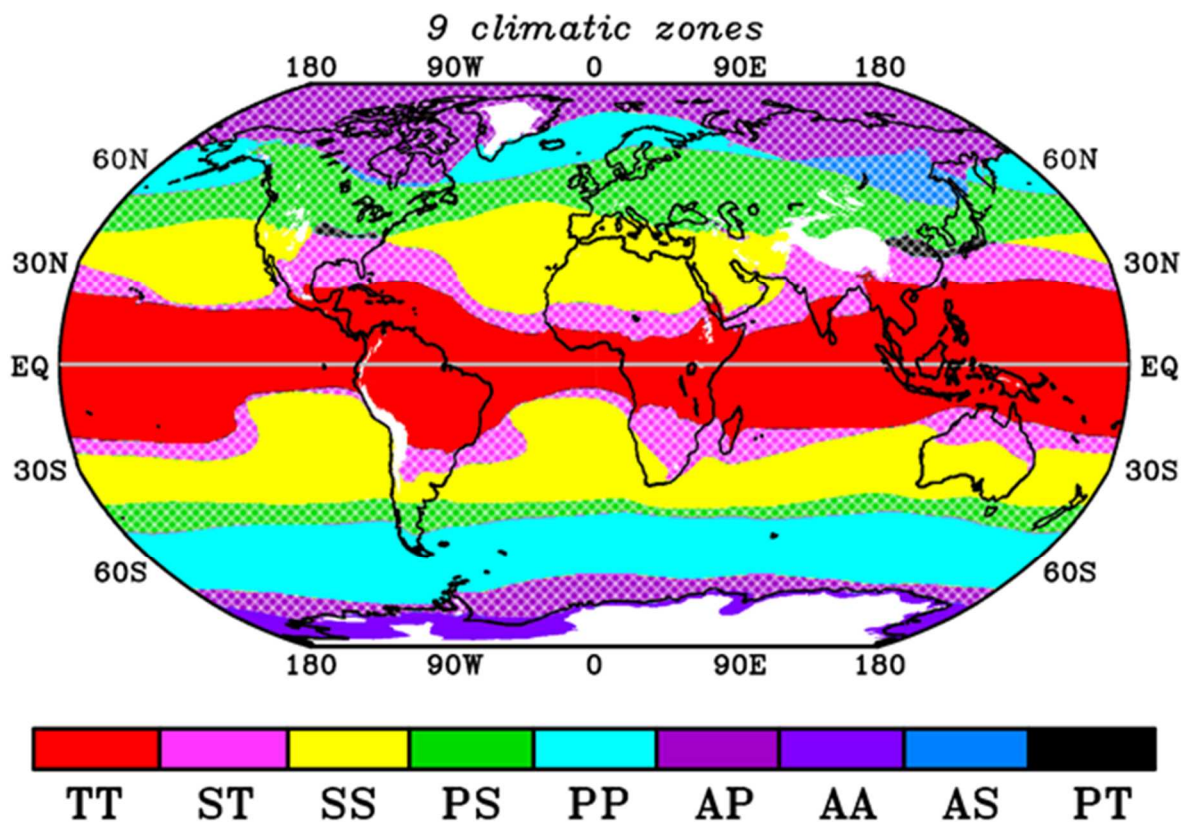


図5 北半球の寒候季（図4a）と暖候季（図4b）における各気団帯の分布域を重ね合わせて得られた9気候帯。ハッチングの領域は、季節により気団帯が入れ替わる地域を表す。空白域は図4と同様。

比較して気候帯が2つ増えた理由は、揚子江流域～西日本や北米大陸の東岸（PT）、シベリア北東部（AS）の地域（図5）において、地理的に隣接する気団帯の単純な入れ替わりではなく、1つとばした気団帯との入れ替わりであることによります（図4 a, bを参照）。これらの地域は、1年のうち気候の変化が大きい地域であると考えられます。

表1 9気候帯の特徴

記号	気候帯の名称	寒候季／暖候季の優勢気団帯	気団帯の挙動
TT	Tropical	Tropical / Tropical	Stable
ST	Quasi-tropical	Subtropical / Tropical	one-class change
SS	Subtropical	Subtropical / Subtropical	Stable
PS	Temperate	Polar / Subtropical	one-class change
PP	Polar	Polar / Polar	Stable
AP	Subarctic	Arctic / Polar	one-class change
AA	Arctic /Antarctic	Arctic / Arctic	Stable
AS	Subarctic	Arctic / Subtropical	two-class change
PT	Temperate	Polar / Tropical	two-class change

3.3. 9気候帯から27気候区への細分化

地球上の気候は、先ず表面太陽放射フラックスの緯度変化の影響を強く受けます（図2）。よって単純にクラスタ数を増やすだけでは、多数の帯状気団に細分されることが予想されます。そこで本研究では、気団の海陸分布を強調するために、帯状平均からのずれ（偏差成分）に着目します。先の解析で同一気団帯にグループ化された格子点の位置に対応する偏差データを用い、各気団帯内を相対的な湿度特性に基づいて乾燥気団(dry)と湿潤気団(moist)に分類しました（詳細は原著論文を参照下さい）。特に顕著な結果を示したのは亜熱帯乾燥気団であり、亜熱帯湿潤気団よりも約12 K気温が高く、約1.3 g kg⁻¹水蒸気量が少ないという特徴がありました。この気団の分布域は世界中の砂漠地域とほぼ一致し、亜熱帯高気圧中心部で下降気流による断熱圧縮によって形成された高温乾燥の空気が、年間を通して支配することでこれらの地域に砂漠化をもたらしていると考えられます。さらに、熱帯湿潤気団は、熱帯乾燥気団と比べて気温はほぼ変わらないが、水蒸気量を約1 g kg⁻¹多く含んでおり、その領域において活発な上場気流と熱帯の対流性降水が起きていると考えられます（図示せず）。

図6は、2季節の乾燥・湿潤気団の分布を重ね合わせて、9気候帯から27気候区へ細分化した図を表します。東西方向の気候差異が考慮された図6は、従来のアリソフの気候区分図（図1）と比較し、大陸性／海洋性気候、大陸上の東岸／西岸気候など、我々の感覚に近いより現実的な気候を再現していま

す。また、図6をケッペン・ガイガーの気候区分図（Köppen and Geiger 1936; Beck et al. 2018）と比較しても、良い対応関係を示すことを確認しました。ここでの純粋な気候データそのものに内在する構造、すなわち気団と前線に基づいて分類した結果は、植生の不連続線を指標にして分類した結果とおおよそ一致している点は、気象学的にも生物地理学的にも興味を惹く結果です。

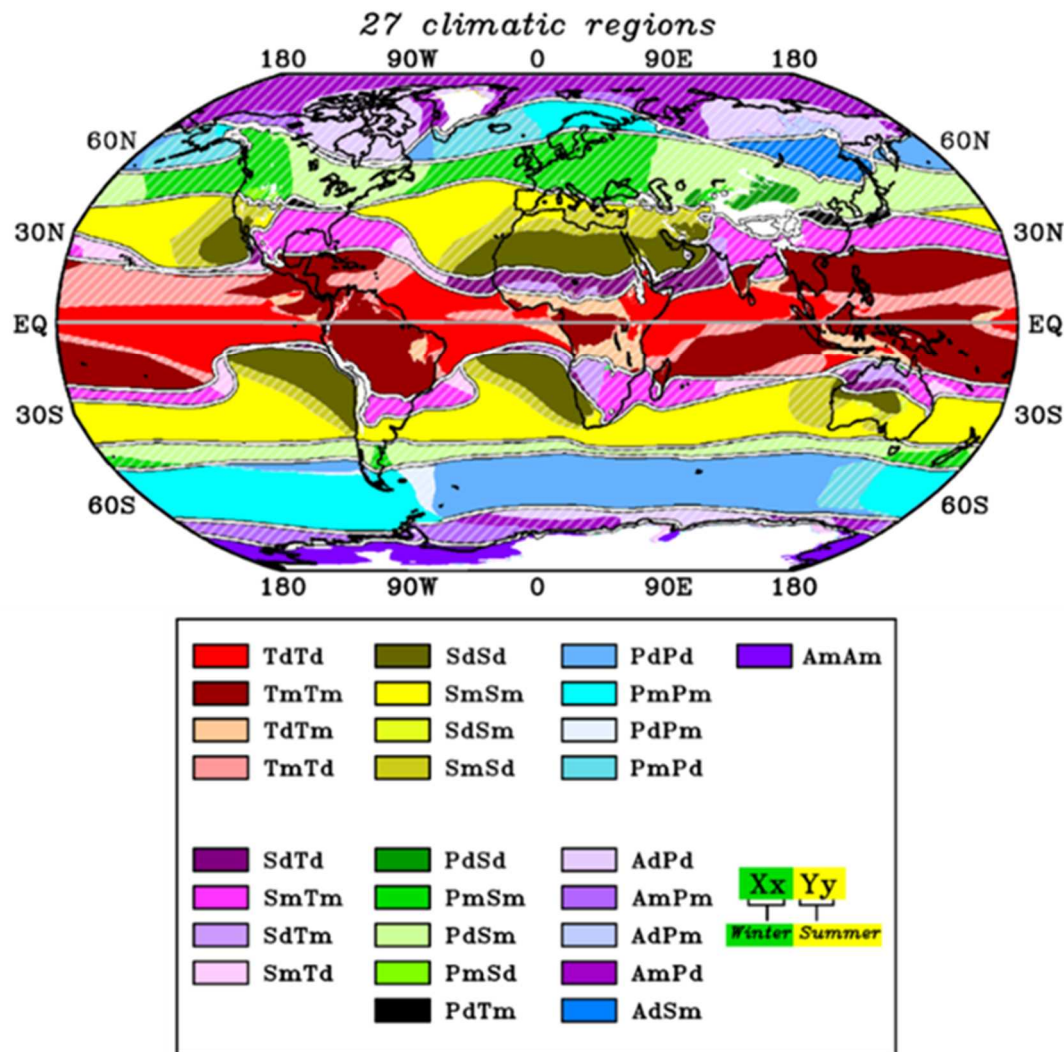


図6 各気団帯内を2クラスタに分類した後、各半球における寒候季と暖候季の気団域を重ね合わせて得られた27気候区（記号の詳細は原著論文をご参照下さい）。ハッチングの領域は、季節により気団が入れ替わる領域を示す。白い等値線は図5の気候帯の境界線を示す。空白域は図4と同様。

4. まとめ

本研究は気団気候学において長年、重要かつ難解な課題の1つとして残されてきた、気団域を客観的に決定する手法を開発しました。すなわち本研究では、全球の再解析データにK-means++クラスタリングを適用し、気団域の定量的かつ客観的な決定を行いました。また、大陸性／海洋性気候、大陸上

での東岸／西岸気候など、ある気候帯内での東西方向の気候差異に基づいて1気候帯内を細分化し、アリソフが1954年に提案するだけにとどまっていた細区分を図示化することに成功しました。

本研究は先ず、1950年代にアリソフが行ったように、地球の気候を4つの気団帯に分けることが、本来は連続的な気候の変化に対してその不連続性を捉えられているのか、という点に疑いを持つことから出発し、4つの定量的な評価指標（DBI、SC、CHI、BIC）を用いてその是非を総合的に判断しました。その結果、地球の気候は4つの気団帯に分けることが統計的に望ましいことが確認されました（図3）。次にクラスタ数（気団帯の数）を4に設定し、クラスタリングで寒候季および暖候季の気団帯の分布域を決定し（図4）、9気候帯を設定しました（図5）。ここでのクラスタリング技術は、中・高緯度の傾圧性に伴う前線性の降水をうまく捉えていることも確認しました。さらに、9気候帯から27気候区への細分化（図6）は、ケッペン・ガイガーの気候区分とも良い対応関係を示すことを確認しました。今後の展望としては、地球温暖化に伴う各気団域と気候学的前線の位置の長期変化を診断することが挙げられます。

参考文献

- Adler, R. F., M. R. P. Sapiiano, G. J. Huffman, J.J. Wang, G. Gu, D. Bolvin, L. Chiu, U. Schneider, A. Becker, E. Nelkin, P. Xie, R. Ferraro, D. B. Shin, 2018: The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, **9**, 138.
- Alisov, B. P, 1954: Die Klimate der Erde (ohne das Gebiet der UdSSR). Deutscher Verlag der Wissenschaften, Berlin, 277pp.
- Arthur, D., S. Vassilvitskii, 2007: k-means++: The advantages of careful seeding. *Proc Symp Discrete Algorithms* 1027–1035.
- Beck, H. E., N.E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, E. F. Wood, 2018: Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci Data*, **5**, 180214.
- Caliński, T. and J. Harabasz, 1974: A dendrite method for cluster analysis. *Commun Stat - Theory and Methods*, **3**, 1–27.
- Davies, D. L. and D.W. Bouldin, 1979: A cluster separation measure. In: *IEEE Trans Pattern Anal Machine Intell*, vol PAMI-1. No. 2, 224–227.
- Flohn, H. 1957: Zur Frage der Einteilung der Klimazonen (The Problem of a Classification of Climatic Zones). *Erdkunde*, **11**, 161–175.
- Hersbach, H., B. Bell, P. Berrisford and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, **146**, 1999–2049.
- Köppen, W. and R. Geiger, 1936: Das geographische system der klimate. In: Köppen W and Geiger R (eds) *Handbuch der Klimatologie*, vol 1. Berlin, Verlag von Gebrüder Borntraeger, 1–44.
- MacQueen, J., 1967: Classification and analysis of multivariate observations. In *5th Berkeley Symp Math Statist Probability*, 281–297.

- 水越 允治・山下 脩二, 1985: 「気候学入門」. 古今書院, 144pp.
- 仁科 淳司, 2019: 「やさしい気候学 第4版: 気候から理解する世界の自然環境」. 古今書院, 144pp.
- Raschka, S. and V. Mirjalili, 2020: Python machine learning programming: Theory and practice by a master data scientist (3rd Edition). Impress, Tokyo, pp 305–311.
- Rousseeuw, P. J., 1987: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.
J. Comput. Appl. Math., **20**, 53-65.
- Schwarz, G., 1978: Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–466.
- 鈴木 秀夫, 1961: 気候区分の諸問題. *地学雑誌*. **70**, 215–219.