

Python を用いた統計的流跡線解析

田崎盛也・城間朝彰・野波秀行・前田美奈子

Statistical Analysis of Trajectories with Python

Moriya TASAKI, Tomoaki SHIROMA, Hideyuki NOHA and Minako MAEDA

要旨：大気汚染物質の発生源の推定には流跡線を利用する方法が一般的である。その中でも多数の流跡線を用いて発生源を推定する方法として CWT (Concentration Weighted Trajectory, 濃度重みづけ流跡線) 解析や、PSCF (Potential Source Contribution Function, 潜在的発生源寄与関数) 解析といった手法がある。本報告では PySPLIT で作成した流跡線データを利用し、Python 上で Ox の解析用データセットの作成と CWT 解析・PSCF 解析を行った。重みづけを行った WCWT・WPSCF 解析からは、中国東部を中心とした地域からの寄与が大きく、Ox 汚染源は主に大陸由来であると推測された。また、元の解析用データセットから、季節など特定の期間や任意の値をとるデータを抽出して解析を行うことも可能であり、Python を用いた統計的流跡線解析は大気汚染物質の発生源推定に有用であると思われる。

Key words: 沖縄県, 統計的流跡線解析, CWT (濃度重みづけ流跡線) 解析, PSCF (潜在的発生源寄与関数) 解析, Python, オキシダント (Ox)

I はじめに

流跡線解析は大気環境の分野ではよく利用されており、ある大気汚染物質の高濃度事象がみられた際にその汚染源地域の推定といった形で用いられることが多い。流跡線の利用には個別の高濃度事象の汚染源推定だけでなく、通年のような長期間において観測点の大気汚染濃度に影響を与える地域を推定するという目的で用いられることもあり、そのような目的に用いられる手法として CWT (Concentration Weighted Trajectory, 濃度重みづけ流跡線) 解析や、PSCF (Potential Source Contribution Function, 潜在的発生源寄与関数) 解析といった多数の流跡線に基づく統計的流跡線解析という手法¹⁾がある。このような手法を行うためには流跡線を多数作成する必要があるが、昨年度の報告²⁾で多数の流跡線を効率的に作成する手法を確立させているため、これを利用し Python で CWT 解析および PSCF 解析を行ったので、その結果を報告する。

II 方法

1. 統計的流跡線の手法について

(1) CWT 解析

CWT 解析は任意の領域を適当な大きさのグリッドセルに分割し、観測点の汚染物質濃度をエンドポイント (計算した遡行時間ごとの位置) 数に応じて加重平均する方法である。各グリッドセルに対して次の式により計算を行う。

$$CWT_{ij} = \frac{\sum_{l=1}^L \sum_{k=1}^K c_l \tau_{ijkl}}{\sum_{l=1}^L \sum_{k=1}^K \tau_{ijkl}}$$

ここで i, j はグリッドセルのインデックス、 k は出発高度のインデックス、 l は流跡線のインデックス、 K は出発高度の数、 L は全流跡線数、 c_l は流跡線 l の観測点濃度、 τ_{ijkl} はグリッドセル ij 内にある出発高度 k 、流跡線 l のエンドポイント数である。CWT_{ij} は空気塊がグリッドセル ij を通る時に観測される汚染物質の平均濃度である。

(2) PSCF 解析

PSCF 解析も同様に次の式で計算される。

$$PSCF_{ij} = \frac{\sum_{l=1}^L \sum_{k=1}^K p_l \tau_{ijkl}}{\sum_{l=1}^L \sum_{k=1}^K \tau_{ijkl}} = \frac{m_{ij}}{n_{ij}}, \quad p_l = \begin{cases} 1 & c_l \geq C \\ 0 & c_l < C \end{cases}$$

p_l は寄与関数であり、 C で設定する閾値以上で 1、閾値未満で 0 の値をとる。 m_{ij} はグリッドセル ij 内にある観測点の汚染濃度が閾値超の全流跡線のエンドポイント数、 n_{ij} はグリッドセル ij 内にある全流跡線のエンドポイント数である。PSCF_{ij} は空気塊がグリッドセル ij を通る時に観測点の汚染物質濃度が閾値以上となる確率である。

(3) 重みづけ

CWT 解析や PSCF 解析の結果はセルの n_{ij} の値が大きいと高い信頼性を持つが、一方で n_{ij} が小さいセルは値の信頼性が低くなるため、セル内の流跡線数によって補正 (Weighted) を行う方がよいとされている。今回の重みづけは Polissir³⁾ や工藤⁴⁾ の用いた係数により補正を行った。

$$W_{ij} = \begin{cases} 1.00, & n_{ij} > 3 \cdot Ave. \\ 0.70, & Ave. < n_{ij} \leq 3 \cdot Ave. \\ 0.42, & 0.5 \cdot Ave. < n_{ij} \leq Ave. \\ 0.17, & n_{ij} \leq 0.5 \cdot Ave. \end{cases}$$

Aveは全グリッド内のエンドポイント数÷総グリッドセル数の値である。

2. 解析に用いたデータと対象地点

解析に用いるデータについて、地点は沖縄局（中部保健所、沖縄県沖縄市）のオキシダント（Ox），期間は2012年度から2021年度までの10年分の1時間値を対象とした。

3. 流跡線解析の設定と解析環境について

流跡線解析の開始時間は測定期間内の毎時、流跡線進行時間は5日間（120時間）とし、気象データはGDASを用い、出発高度は500 m, 1000 m, 1500 mの3つとした。流跡線ファイルの作成やその後の統計的流跡線解析にはAnaconda ディストリビューション (<https://www.anaconda.com>)により解析環境を構築し、プログラミング言語 Python (3.7.12) と PySPLIT (0.3.6) を使用した。その他のライブラリについては表1に示す。なお、Anaconda ディストリビューションを使用せず公式版 Python (<https://www.python.org>)を利用して環境構築を行い同様にコードが動作することも確認している。またPySPLITの使用にはデスクトップ版のHYSPLITが必要となるためこれも導入しておく必要がある。流跡線作成のコードはコード1を参照（Pythonコードはこの報告の付録を参照）。コード1の実行により262,944個の流跡線ファイルを作成した。流跡線作成はIntel Core i9-11900搭載のPCで1日と10時間程度の実行時間が必要となる。解析した流跡線ファイルのうち、逆解析流跡線による相対誤差が基準を超えるものや解析範囲の高度条件を逸脱するものなどについては解析対象から除外することとした。これにより26,886個（10.2%）の流跡線ファイルを除外し、236,058個の流跡線を有効とした。

表1. 使用した Python と主なライブラリのバージョン。

Name	Anaconda	公式版 Python
	Version	Version
Python	3.7.12	3.10.9
Numpy	1.21.5	1.24.4
Xarray	0.20.2	2023.6.0
pandas	1.3.5	2.0.3
GeoPandas	0.10.2	0.13.2
Shapely	1.8.0	1.8.2
Matplotlib	3.5.1	3.7.2
japanize-matplotlib	1.1.3	1.1.3
Basemap	1.2.2	1.3.7
PySPLIT	0.3.6	0.3.6

4. 解析用データセットの作成について

流跡線データを用いて統計的流跡線解析を行う場合、緯度・経度や観測点の時間、流跡線の遡行時間など多次元のデータになり通常の表形式のデータでは扱いが難しい点があるため Python の Xarray というライブラリを使用し、解析に必要な情報を含んだ解析用データセットを作成した。データセットの作成については付録のコード6,7を参照。

5. 統計的流跡線解析の設定について

CWT 解析, PSCF 解析いずれについてもグリッドセルの範囲は 0–50N°, 90–160E° とし, 1° × 1° のサイズで分割した。PSCF 解析では閾値を設定する必要があるが, Ox の環境基準 60ppb を閾値として使用し, PSCF 解析を行った。

III 結果および考察

1. エンドポイントの頻度と重みづけについて

統計的流跡線解析を行うにあたって、グリッドセル上のエンドポイントの頻度や重みづけを確認することも重要である。図1にグリッドセル上のエンドポイントの頻

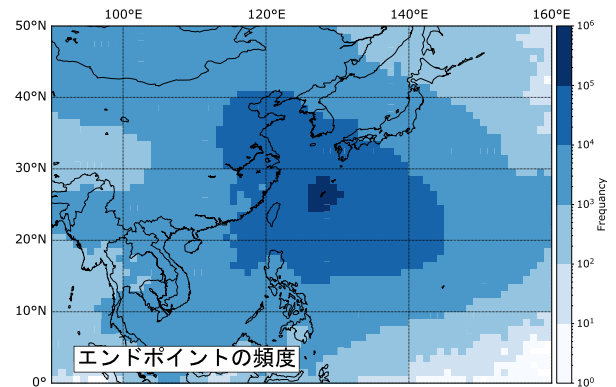


図1. 流跡線のエンドポイントの頻度分布。

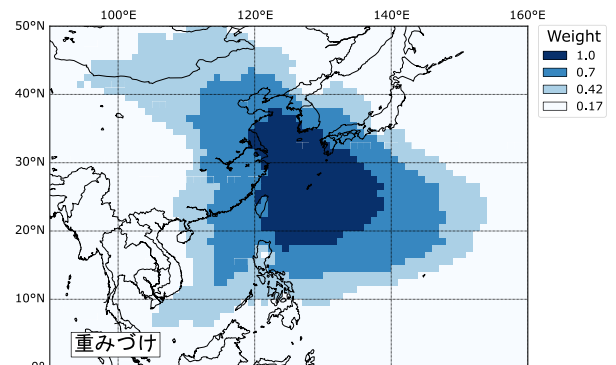


図2. 重みづけの分布。

度を 10 の N 乗毎に階級表示した分布を示す。観測点のある沖縄島付近にエンドポイントが集中し、観測点から離れるにしたがって頻度が減少しているのが分かる。これを反映し、沖縄近海や中国東部の沿岸部などが重みづけ係数 1 をとる領域となっている (図 2)。統計的流跡線解析の結果に重みづけを適用する場合、この重みづけの妥当性や、そもそも適用する対象物質が移流を想定できるような物質であるかということを検討しておく必要がある。

2. CWT および WCWT 解析

CWT および WCWT 解析の結果を図 3 上段に示す。CWT 解析では沖縄島よりも北の広い範囲に発生源があるように見える。これは Ox 濃度が高い春季・冬季は主に大陸等から、Ox 濃度が低い夏季は主に太平洋からの空気塊流入があることを反映している。エンドポイント頻度で重みづけをした WCWT では中国東部・黄海・東シナ海近辺からの影響が大きいことが示唆される。

3. PSCF および WPSCF 解析

PSCF および WPSCF 解析の結果を図 3 下段に示す。PSCF 解析では中国東部や東北部、内陸部やモンゴルまで発生源があるように見えるが、流跡線の数で重みづけ

をした WPSCF では中国東部からの影響が大きいことが示唆される。

4. 季節ごとの結果について

Xarray でデータセットを作成することによって、元のデータセットから期間や項目を任意の値に設定したデータを抽出することができる。ここでは任意の期間として季節を対象とし、夏季と冬季の結果について紹介する。

(1) 夏季 (6-8 月)

夏季の WCWT および WPSCF 解析の結果を図 4 上段に示す。夏季は流跡線の多くが太平洋など南から由来し、大陸から通過してくるものはわずかなため、重みづけの結果 WCWT 解析では東シナ海の WCWT 値が高くなる結果となっている。また、WPSCF 解析では重みづけがされても大陸からの影響による高濃度事象があることが確認できる。

(2) 冬季 (12-2 月)

冬季の WCWT および WPSCF 解析の結果を図 4 下段に示す。冬季は流跡線の多くが大陸方向から到来しているため、重みづけの結果 WCWT 解析では全データの際よりも内陸部の影響度が増している。WPSCF 解析の結果では、中国東部を中心とした地域からの移流が見られることがわかる。

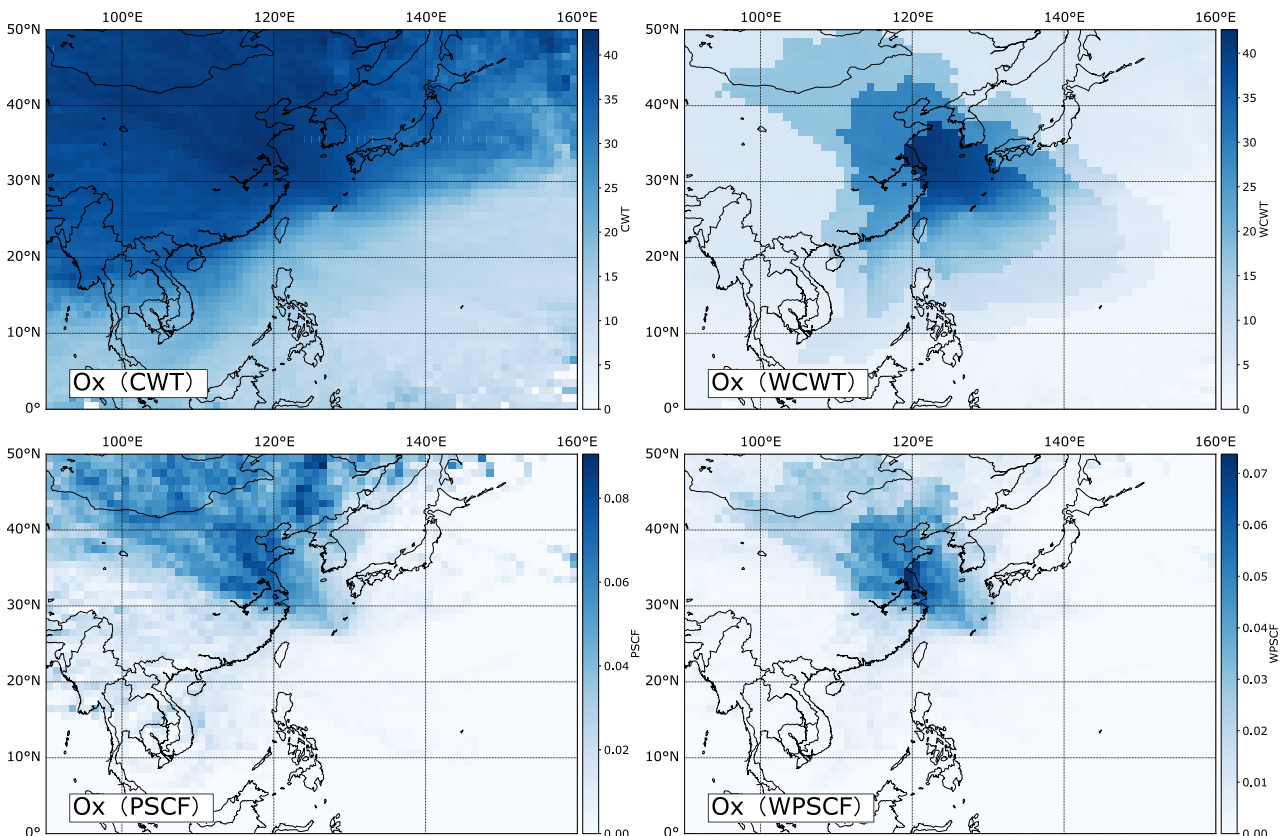


図 3. Ox の統計的流跡線の解析結果.

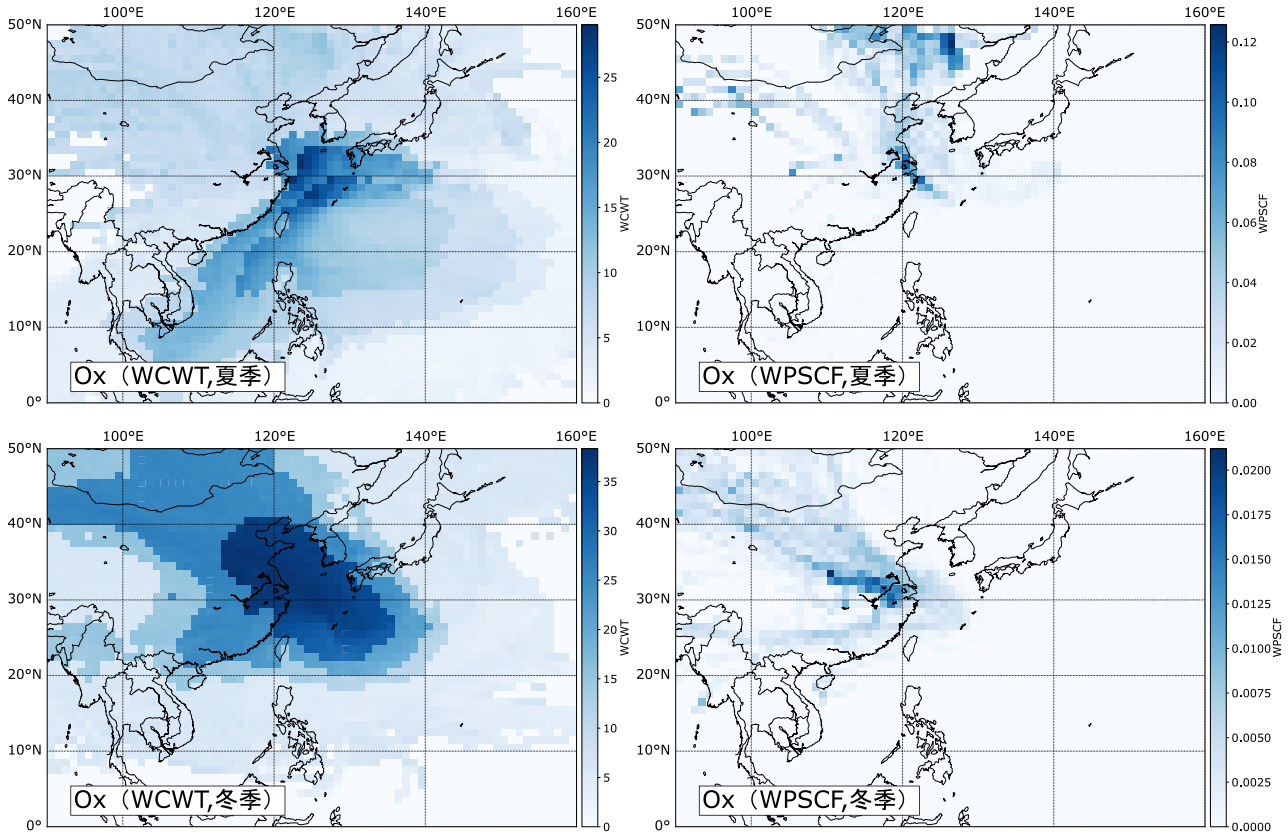


図4. Oxの夏季および冬季のWCWT, WPSCF解析結果.

IV まとめ

統計的流跡線解析の手法であるCWTおよびPSCF解析について、プログラミング言語Python上で解析を可能とした。それにより多数・長期間（今回の例では10年分）のデータに対する解析の効率的実行が可能となった。また、本報告で解析した結果によりOxの汚染源が主に大陸由来であると推測され、エンドポイント頻度の分布から空気塊移流の特徴も明らかとなった。さらに、季節ごとの結果などXarrayライブラリ利用で任意のデータ抽出も容易であり、Python上で統計的流跡線の解析を行うことは有用であると考えられる。

V 参考文献

1) 大気環境学会編 (2019) 大気環境の事典. 朝倉書店,

pp.112-113.

2) 田崎盛也 (2022) Pythonによる流跡線解析. 沖縄県衛生環境研究所報, 56 : 78-101.
 3) Polissar, A. V., Hopke, P. K., Harris, J. M (2001) Source regions for atmospheric aerosol measured at Barrow, Alaska, Environ. Sci. Technol., 35, 4214-4226.
 4) 工藤慎治・飯島明宏 (2017) クラスタ分析およびCWT解析を用いた東京のPM_{2.5}質量濃度の増加に影響を及ぼす地域の推定. 大気環境学会誌, 52 : 89-99.

VI 付録

本報告に関連するコードについて、衛生環境研究所ホームページ (<https://www.pref.okinawa.jp/site/hoken/eiken/kankyo/taiki.html>) にて掲載する。