

Support Vector Machines を用いた極成層圏雲の表面積の推定

池田 奈生[†] 芦田 尚美[‡] 林田 佐智子[‡]

[†] 奈良女子大学大学院 人間文化研究科

[‡] 奈良女子大学 理学部情報科学科

オゾンホールとして広く知られている化学的オゾン破壊は、冬季北極域の高度約 20 km 付近で起こる。この化学的オゾン破壊量は、気温が約-78 度以下になると発生する極成層圏雲 (Polar Stratospheric Clouds : PSC) の表面積に大きく依存する。しかし、PSC の観測例は少なく、また数値モデルを用いた推定では、気温の変動が PSC 推定に与える影響が大きいため、正確な推定が難しい。

そこで本研究では、衛星観測データと光化学ボックスモデルを用いて、PSC 表面積を推定する。そして、この推定結果をサポートベクターマシンに適用し、PSC 表面積推定システムを構築する。

Estimation of the surface area density of Polar Stratospheric Clouds by using the Support Vector Machines

NAO IKEDA[†] NAOMI ASHIDA[‡] SACHIKO HAYASHIDA[‡]

[†] Graduate School of Human Culture, Nara Women's University

[‡] Department of Information & Computer Sciences, Nara Women's University

Chemical ozone destruction occurs at around 20 km altitude in the Arctic winter, and the amount of ozone loss depends on the surface area density of Polar Stratospheric Clouds(PSC) which appear at the temperature below -78 °C. However, there are not enough observations for PSCs, and it is difficult to estimate PSC surface on numerical simulations because the change of the temperature leads the large error in the estimation.

In this study, we estimate the surface area density of PSC using satellite data and a chemical box model. Then we build a estimation system of the PSC surface area by using the Support Vector Machines.

1 はじめに

冬季極域での大規模な化学的オゾン破壊は、一般にオゾンホール現象として知られている。これは冬季極域に発生する PSC 上で、オゾン破壊しない ClONO₂ や HCl といった不活性塩素が、オゾン破壊する Cl や ClO といった活性塩素に多量に変換され、オゾンと反応するためである。したがって、PSC はオゾン破壊に大きな影響を及ぼす。

PSC 上での反応は PSC の粒子表面積に比例する。したがって PSC の表面積は化学的オゾン破壊量を理解する上で重要なパラメータである。しかしながら冬季北極では、年ごとに PSC 発生量が大きく異なり、さらに PSC 表面積の観測事例も少ないため、見積もりが難しい。

そこで本研究では、PSC に大気中の存在量が大きく依存する化学種である ClONO₂ の観測値と、ClONO₂ に関連するパラメータから PSC 表面積を推定するシステムの構築を目的とする。そこで、このシステム構築の第 1 歩として、光化学ボックスモデルを用いて

ClONO₂ 観測値を再現し、この結果をサポートベクターマシン (Support Vector Machines : SVM) に適応させる。SVM は Vapnik ら [7, 6] によって提案された 2 クラス分類器で、少ない学習データで特徴の次元を増やすことが可能であり、本研究で行うようなシステムの構築に適している。以上のように構築されたシステムについての評価・検討をおこなう。

2 学習データセット作成

ここでは SVM に適応させるデータ作成について説明する。

2.1 ClONO₂ 観測データ

1996 年に打ち上げられた日本の衛星「みどり」に搭載された改良型大気周縁赤外分光計 (Improved Limb Atmospheric Spectrometer : ILAS) は、世界で初めて冬季北極域での ClONO₂ の長期間連続観測を行った [4]。大気中の ClONO₂ 量は PSC の発生に大きく依存するため、ClONO₂ 量の推定から PSC 表面積推定が可能となる。

本研究では ILAS の最新の Version 7 のデータを使

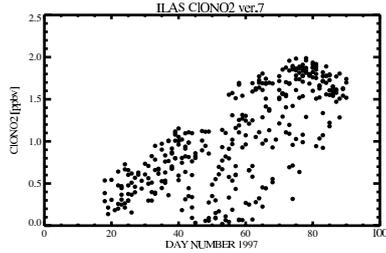


図1 ILASが観測した温位475K, 極渦内での1997年1-3月のClONO₂の観測値。縦軸はClONO₂[ppbv], 横軸は1997年1月1日を1とした日数をそれぞれ表す

用する。このデータでは、これまで導出が困難であったPSC発生時のデータも扱うことができる[5]。この観測データの中から北極域の1月から3月、オゾン破壊量が最も多い温位475K(高度約20km)の極渦内のデータを使用する。温位とは、ある気圧、気温状態での空気を断熱的に地表面気圧まで持ってきたときの温度のことである。一般に温位は位置・高度をかえても保存されるため、空気は同じ温位面上を移動すると考えてよい。使用するClONO₂観測値を図1に示す。ClONO₂は季節の進行とともに増加傾向を示すが、PSC上での不均一反応が起こると減少する。図1では20日付近と40-60日付近で不均一反応がおこっている。また、ClONO₂はばらついた値をとるが、これは観測値ごとに、観測される以前に経験した不均一反応の規模が異なるためである。本研究ではこの観測値を使用する。

次にClONO₂観測値を再現するために使用する2つのモデルについて説明する。まず1点目は流跡線計算モデルである。これは観測された空気塊が移動した位置を計算するモデルである。ILASは大気中の空気塊を観測しており、この空気塊は、空気塊ごとに異なる軌跡により輸送されたものである。したがって空気塊がこれまでに経験した光化学反応を再現するためには、観測値ごとに空気塊が移動してきた場所を推定し、その位置で光化学ボックスモデルを計算させなければならない。そこで本研究では空気塊の輸送経路を推定する流跡線計算を行うために、宇宙航空研究開発機構/地球観測研究センターが開発した流跡線計算モデル、EORC-TAM(Earth Observation Research Center Trajectory Analysis Model)[3]を使用する。このモデルでは入力に初期地点の緯度、経度、高度を与えると、指定時間までの緯度、経度、高度、気温を1時間ごとに出力する。ここで使用されている気象場のデータはEuropean Centre for Medium-Range Weather Forecasts(ECMWF)の全球客観解析データである。また本研究で行う計算方法は475K温位面にそって観測日時から7日前までさかのぼる後方流跡線計算である。

次に、光化学ボックスモデルについて説明する。光

化学ボックスモデルは大気中の化学種の濃度の時間変化を計算するモデルである。本研究で使用するモデルは入力に各化学種の初期値、緯度、経度、高度の時間変化を与え、指定時間後の各化学種の濃度を出力するものである。本研究ではKagawa and Hayashida[2]において使用されたモデルを基礎として構築する。このモデルは成層圏の化学反応を表現できるように59種類の化学種、101種の気相反応、48種の光解離反応、7種の不均一反応を含む。

次にデータセットの作成方法について説明する。データセットは以下の手順で作成する。

1. EORC-TAMを用いて、すべてのILAS観測地点より温位475Kで、観測日時から観測7日前までの流跡線を計算する
2. 観測の7日前から観測日まで、EORC-TAMの出力位置情報にそって光化学ボックスモデル計算を行う。このとき気温バイアスを考慮してEORC-TAMで出力された気温履歴と、この気温履歴プラスマイナス5℃の間を1℃刻みで計11通り計算する。1997年の1月1日から3月31日まで、極渦内では計311点の観測結果が存在する。したがって、311(観測地点数)×11(気温履歴)の計3421本のモデル計算を行う
3. 観測イベントごとに、ClONO₂観測値ともっとも近いモデル計算結果を選択する。
4. 選択したモデル計算結果から、ClONO₂に関連する以下の7つのパラメータを用いてデータセットを作成する。
選択したパラメータは
(a) ClONO₂観測値
(b) 7日間のPSC表面積の合計
(c) 1997年1月1日を1とする日数
(d) 7日間の日照時間の合計
(e) 7日間のClO量の合計
(f) 7日間のNO₂量の合計
(g) ECMWFの気温のバイアスの7つである。

3 PSC 表面積推定システムの構築と検証

3.1 サポートベクターマシン

本研究で行うPSC表面積の推定は有限区間を分割した、多クラス分類問題に帰着することができる。そこで多クラス分類法として、SVMを用いる。SVMは多次元の特徴ベクトルを、2つのクラスに識別する2値クラスの線形識別器である。識別器は $f(x) = w \cdot x + b$ で表され、 $f(x)$ の値によってクラスに分類する。ここで w は重みベクトル、 x は入力ベクトル、 b はバイアスである。 $f(x) = 0$ を満たす面を超平面と呼ぶ。これを複数組み合わせることにより、多クラス分類問題への応用が可能である。また、SVMの特徴としてマージン最大化があげられる。マージンとは超平面と特徴ベクトル間の最小距離である。このマージンを最大化す

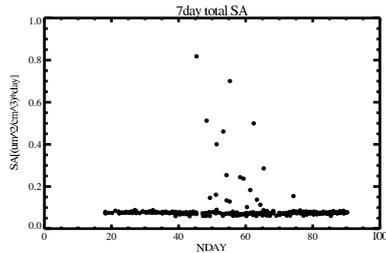


図2 7日間のPSC表面積積分値. 横軸が1997年1月1日を1とした日数, 縦軸がPSC表面積の積分値をそれぞれ表す

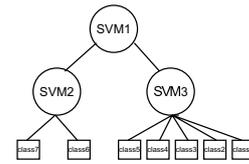


図3 階層的SVMの概略図

ることによって学習データの識別精度を高めている。これにより、学習データが少なくても特徴ベクトルの次元を増加させることが可能である。さらに、非線形問題への対応はカーネル関数を用いて実現されている。本稿では、LIVSVM[1]を用いてSVMの実装を行う。LIVSVMは台湾国立大学のLinらによって作成されたSVMのライブラリで、シンプルで使いやすいツールとして提供されている。カーネル関数は、クラスと属性の関係が非線形でも対応可能なRadial Basis Function(RBF)を使用する。

3.2 PSC表面積データ

本研究でクラス分類するパラメータはPSC表面積である。図2に、作成したデータセットのPSC表面積の値を示す。この図は、観測された空気塊が過去7日間にどれだけのPSC表面積を経験したかを示している。図2より大部分のデータが0.1以下であることがわかる。ここで、0.1以上の値はオゾン破壊に大きく影響を及ぼすPSCであることを示している。また0.1以下のデータはPSCではない、バックグラウンドに存在する粒子の表面積か、あるいはオゾン破壊に少しだけ影響を及ぼしたPSCである。観測値の合計311点のうち、PSCの表面積が0.1以上は19点、0.1未満の点は292点であった。

3.3 PSC表面積システムの構築

PSC表面積データの分布は3.2節で示したように表面積の値が0.1を境に明らかに特徴が異なるため、階層的なシステムの構築を試みた。図3にこのシステムの概略図を示す。

まずPSC表面積が0.1以上のクラスと0.1未満のクラスに分類するSVMを作成した。このSVMを「SVM1」とする。次にPSC表面積が0.1以上のクラスに分類されたデータをさらに分類するSVMを作成した。このSVMを「SVM2」とする。ここでのクラス数はデータが少量であることから、クラス分類が可能かどうかを検討するために2クラス(クラス6,7)での分類を試みた。一方、PSC表面積が0.1以下のデータは300点近く存在する。PSC表面積の有効桁数の関係より、データが存在する範囲を等間隔で5クラス(クラス1-5)に分類することを試みた。このSVMを

「SVM3」とする。表1にSVM2とSVM3の、クラス分類時のPSC表面積の範囲を示す。

次に、これらのSVMを作成するにあたり次の条件で実験を行った。

1. 311個の全データのうち、トレーニングデータとテストデータの個数を200:111(実験1)又は160:151(実験2)にわけて実験を行った。
2. 各実験では、異なった分け方で10種類のトレーニングデータ集合・テストデータ集合のペアを作成し、10回の実験を行った。

表2に実験1の結果を示す。この結果より、全体を2クラスに分類するSVM1では10回の実験の平均値が98.47%と高い正答率での判定が可能であった。次に、PSC表面積が0.1以上の部分を2クラスに分類するSVM2では10回の実験の平均正答率が50.67%となり、識別が難しいことがわかった。一方、PSC表面積0.1以下を5クラスに分類するSVM3では、10回の実験の平均正答率が80.48%となった。したがってSVM1とSVM3を用いた階層的なPSC表面積推定システムは、構築が可能であることがわかった。

実験2の結果を表3に示す。結果よりSVM1では98.08%という高い正答率を示したが、SVM2とSVM3では正答率が80%に満たず、トレーニングデータが160個ではシステム構築は難しいことがわかった。今回の実験を通して、1回の学習をするのに要した時間は、AMP Athlon Sempron 1.833GHz, メモリ

表1 クラスごとのPSC表面積の範囲

class name	PSC Surface Area(SA)[μ^2/cm^3]
class 1	$SA \leq 0.065$
class 2	$0.065 \leq SA < 0.071$
class 3	$0.071 \leq SA < 0.077$
class 4	$0.077 \leq SA < 0.083$
class 5	$0.083 \leq SA < 0.10$
class 6	$0.10 \leq SA < 0.40$
class 7	$0.40 \leq SA$

256MB の Windows 2000 上で数分程度であった。

4 まとめ

本論文では ILAS が観測した ClONO₂ データと光化学ボックスモデルの結果からデータセットを作成し、これを SVM に適応することによって PSC 表面積推定システムの構築を試みた。システム構築にあたり、PSC 表面積に明らかに異なる 2 つの特徴が見られたため、2 層構造システムの構築を行った。

その結果、PSC 表面積データ分布の明らかな 2 つの特徴については、高い識別精度で識別することが可能であった。このことは、オゾン破壊に大きな影響を及ぼす PSC を高い精度で判別できたことを示しており、大気化学の分野への貢献が可能である。また下層の SVM においてもデータ数が十分に確保できる場合は識別可能であった。しかしながら、データ数が少ない部分では識別が困難であった。従って、本研究で提案した PSC 表面積推定システムは、十分なデータ数が確保できる場合には構築が可能であることがわかった。

今後の課題として、まずデータセットの確度を高めるため、光化学ボックスモデルの初期値の考慮や観測値の誤差の評価の必要性があげられる。また、実用的なシステム構築のため、SVM に与えるパラメータについては、観測値が存在するものを選択するなどの検討が必要である。

謝辞

流跡線ツールを提供してくださった宇宙航空研究開発機構/地球観測研究センターの皆様、ILAS のデータを提供してくださった ILAS サイエンスチームの皆様、特に Version 7 データの提供に対し、国立環境研究所、ILAS チームの皆様にご感謝いたします。TUV は NCAR/ACD で提供いただきました。

表 2 実験 1 の正答率の平均と分散

SVM	mean [%]	deviation [%]
SVM1	98.47	0.82
SVM2	50.67	185.45
SVM3	80.48	2.77

表 3 実験 2 の正答率の平均と分散

SVM	mean [%]	deviation [%]
SVM1	98.08	1.09
SVM2	53.33	242.42
SVM3	77.96	11.31

参考文献

- [1] Chang, C.-C. and Lin, C.-J.: *LIB-SVM: a library for support vector machines* (2001). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [2] Kagawa, A. and Hayashida, S.: Analysis of ozone loss in the Arctic stratosphere during the late winter and spring of 1997 using the Chemical Species Mapping on Trajectories (CSMT) technique, *J. Geophys. Res.*, Vol. 108, No. D22, p. 4698 (2003).
- [3] Matsuzono, T., Sano, T. and Ogawa, T.: Development of the Trajectory Analysis Model(EORC-TAM), *EORC Technical report* (1998).
- [4] Nakajima, H., Sugita, T., Irie, H., Saitoh, N., Kanzawa, H., Oelhaf, G., Wetzell, G., Toon, G. C., Sen, B., Blavier, J.-F., Traub, W. A., Jucks, K., Johnson, D. G., Yokota, T. and Sasano, Y.: Measurements of ClONO₂ by the Improved Limb Atmospheric Spectrometer (ILAS) in high-latitude stratosphere: New products using version 6.1 data processing algorithm, *J. Geophys. Res.*, Vol. 111, No. D11S09 (2006).
- [5] Oshchepkov, S., Sasano, Y., Yokota, T., Nakajima, H., Uemura, N., Saitoh, N., Sugita, T. and Matsuda, H.: ILAS data processing for stratospheric gas and aerosol retrievals with aerosol physical modeling: Methodology and validation of gas retrievals, *J. Geophys. Res.*, Vol. 111, No. D02307 (2006).
- [6] Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer-Verlag (1995).
- [7] Vapnik, V. and Lerner, A.: Pattern recognition using generalized portrait method (1963).