

気象衛星データを用いた 機械学習による可降水量推定手法の検討

1220103 谷村朋哉

高知工科大学 システム工学群 建築・都市デザイン専攻

Email: 220103f@ugs.kochi-tech.ac.jp

気候変動を監視するうえで、最大の温室効果を持つ水蒸気について把握することが重要である。本研究では、気象衛星ひまわりのデータを用いた可降水量推定手法の構築に役立てるために、機械学習手法の一つであるランダムフォレストを用いて可降水量を推定し、観測時期および観測点による比較を行うことで、可降水量の推定における精度検証や重要変数の調査を行った。結果として、位置条件による推定精度の違いに傾向は見られなかった。また、可降水量の推定モデルは各年、各月それぞれで作成される必要があることが分かった。

Key Words: 可降水量, 機械学習, ランダムフォレスト, 気象衛星ひまわり

1. はじめに

1.1 背景

大気中の水蒸気は地表や大気からの赤外放射を吸収し、かつ大気中に豊富に存在する気体であるため、最大の温室効果を持つ。水蒸気は人為的に排出される温室効果ガスには含まれないが、他の温室効果ガスによって引き起こされる温暖化をさらに増幅させると考えられる¹⁾。そのため、地球温暖化等の気候変動を把握するうえで、大気中の水蒸気量は無視できない。大気中の水蒸気量を表す指標として、可降水量が挙げられる。可降水量は、地上の単位面積に立てた鉛直気柱内に含まれる全水蒸気量と定義される。赤塚は、GMS-5 用に開発された赤外スプリットウィンドウチャンネルを用いた可降水量推定手法をMTSAT データに適用し、MTSAT データ用の可降水量推定手法の開発を行った²⁾。次に、赤塚は観測頻度が高い気象衛星ひまわり 8 号による可降水量推定手法の検討を行った¹⁾。ここでは、MSG-SEVIRI による可降水量推定手法と同様の特徴量を用いたランダムフ

ォレスト (RF: Random Forest) 回帰による推定手法を検討したが、今後詳細な検討が必要であるとしている。

1.2 目的

本研究では、観測頻度が高い気象衛星ひまわりから得られた輝度温度等のデータを用いて可降水量を面的に推定する手法について検討を行う。本研究の独自性として、後述する RF の回帰分析により推定モデルを作成する。そして、月毎および観測点毎に推定モデルの精度検証や重要変数の検討を行うことで、推定モデルの精度向上のための知見を得ることを目的とする。

2. 手法

2.1 可降水量推定手法

可降水量推定手法として機械学習手法の 1 つである RF を用いた。RF は、複数のアルゴリズムを組み合わせることにより予測性能を向上させるアンサン

表-1 説明変数

① 緯度	Lat
② 標高	Ele
③ 衛星天頂角	VZA
④ 観測バンドの輝度温度	(例) B08, B09
⑤ 輝度温度差	(例) B09-B08, B10-B08
⑥ 輝度温度差の2乗	(例) $(B09-B08)^2$
⑦ 輝度温度の鉛直成分	(例) $B08\cos VZA$
⑧ 輝度温度差の鉛直成分	(例) $(B09-B08)\cos VZA$
⑨ ⑥の鉛直成分	(例) $((B09-B08)^2)\cos VZA$
⑩ 輝度温度の自然対数	(例) $\ln(B08)$
⑪ 輝度温度差の自然対数	(例) $\ln(B09-B08)$
⑫ ⑥の自然対数	(例) $\ln((B09-B08)^2)$
⑬ ⑩の鉛直成分	(例) $(\ln(B08))\cos VZA$
⑭ ⑪の鉛直成分	(例) $(\ln(B09-B08))\cos VZA$
⑮ ⑫の鉛直成分	(例) $(\ln((B09-B08)^2))\cos VZA$

ブルアルゴリズムである³⁾。本研究では、予測値の生成に150本の決定木を使用し、目的変数を2016年のラジオゾンデ観測から得られる可降水量、説明変数を各ラジオゾンデ観測点の緯度、標高、2016年のひまわり赤外バンドの輝度温度等(表-1, 表-2)とし、RF回帰によって可降水量を推定する。説明変数の選定には、GMS-5やMTSATを用いた赤外スプリットウィンドウチャンネルを用いた可降水量アルゴリズムを参考にした。この時、2016年1年間のデータのうち80%を訓練データ、20%をテストデータとし、2017年の可降水量データを検証用データとした。このうち、全観測点の2016年訓練データを使用して月毎の可降水量推定モデルを作成する。そして、全観測点の2016年テストデータおよび2017年の検証用データを用いて精度検証を行う。また、その後はPermutation Importanceと呼ばれる測定方法を用いて説明変数の重要度を明らかにする(図-1)。この測定方法では各説明変数をシャッフルし、精度の低下が大きいものを重要な変数として捉えることができ

表-2 ひまわり8号の観測バンド⁴⁾

バンド番号	中心波長(μm)	用途の例
8	6.2	上層水蒸気量
9	6.9	上中層水蒸気量
10	7.3	中層水蒸気量
13	10.4	雲画像、雲頂情報
14	11.2	雲画像、海面水温
15	12.4	雲画像、海面水温

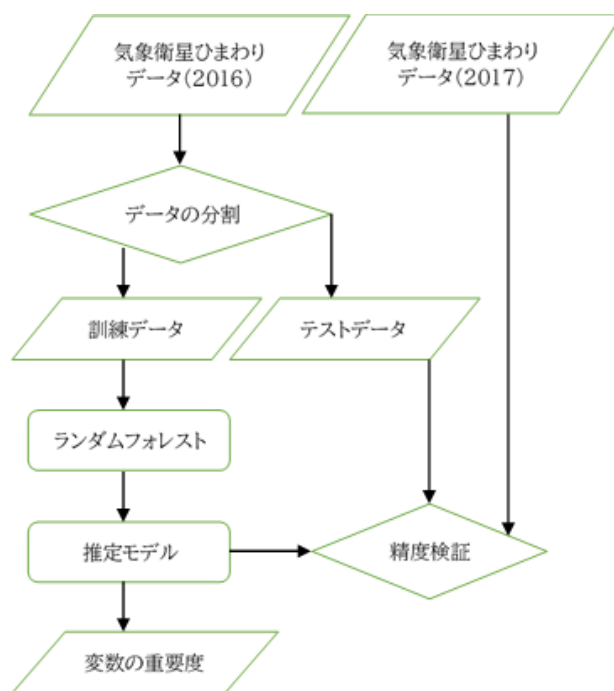


図-1 可降水量推定モデル作成のフローチャート

る⁵⁾。

2.2 使用データ

本研究では、説明変数として気象衛星ひまわりにより得られた輝度温度⁶⁾、目的変数として2016年の各月、各ラジオゾンデ観測点での可降水量⁷⁾を使用した。また、検証には2017年の気象衛星ひまわりデータとラジオゾンデの可降水量データを使用した。

2.3 解析手法

本研究では、可降水量の推定モデルを2種類に分

けて作成した。一つは、2016 年の全観測点訓練データを用いた推定モデルである。もう一つは、2016 年の各観測点訓練データを用いた推定モデルで、後者は Lat, Ele をインプットデータから除いた。

a) 作成モデルの他の年への適用可能性の検討

2016 年全観測点訓練データを用いてモデルを作成し、同年のテストデータおよび 2017 年の検証用データにおける RMSE を計算・比較する。このことから、2016 年データから作成された推定モデルが、他の年のデータへの利用が可能であるか調べる。

b) RMSE の空間的特徴の検討

第一に 2016 年全観測点データを用いてモデルを作成し、各観測点データを入力して RMSE を計算・比較する。このことから、全観測点から作成したモデルを使用する際、各観測点においてどの程度誤差が生じるか調べる。第二に 2016 年各観測点データを用いてモデルを作成し、同様に RMSE を計算する。作成した二つのモデルにおいて、RMSE と VZA および RMSE と Ele の関係を調べることにより、VZA および Ele が可降水量の推定に及ぼす影響を考察する。

c) 月毎の重要変数についての検討

2016 年全観測点データを用いてモデルを作成し、Permutation Importance を用いて説明変数の重要度を計算する。ここで得られた重要度が大きい上位 10 個の変数を重要変数とする。そして、重要変数の月毎の違いについて検討することで、年間を通して 1 つの推定モデルで可降水量の推定が可能であるか調べる。

3. 結果および考察

3.1 可降水量推定モデル

2016 年の訓練データから作成した推定モデルは、2016 年テストデータを用いた精度検証では RMSE が 4.01mm から 7.26mm の値を取った(表-3)。これは赤塚らによる重回帰分析²⁾の解析結果である 6.5mm から 10.5mm の RMSE を下回り、精度の向上が見られた。

3.2 解析結果の比較

a) 作成モデルの他の年への適用

各観測点データを入力した時の 2016 年のテストデ

表-3 全観測点データを用いた解析による 2016 年および 2017 年の各月 RMSE

月	RMSE (2016) [mm]	RMSE (2017) [mm]
1	4.02	6.23
2	4.64	5.26
3	5.26	6.50
4	5.33	7.85
5	6.16	8.23
6	5.84	7.07
7	6.29	7.12
8	7.26	7.61
9	6.06	8.19
10	5.90	7.52
11	4.68	6.51
12	5.28	5.63

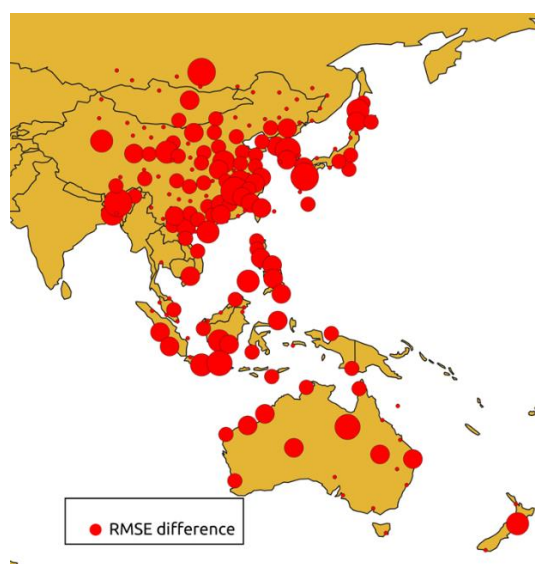


図-2 2016 年 RMSE と 2017 年 RMSE の差
(RMSE difference の大きさは RMSE の差に対応)

ータにおける RMSE と 2017 年検証用データにおける RMSE を比較した(図-2)。両者の差は 0.01 から 8.23 の値をとった。図からも分かるように RMSE の差は観測点によって開きがあるものの位置的な偏りは見られないため、対象とした観測地域では、2016 年データから作成した推定モデルを他の年へ適用することは困難であると考えられる。また、全観測点を対象とした時の 2017 年各月の RMSE を表-3 に示す。

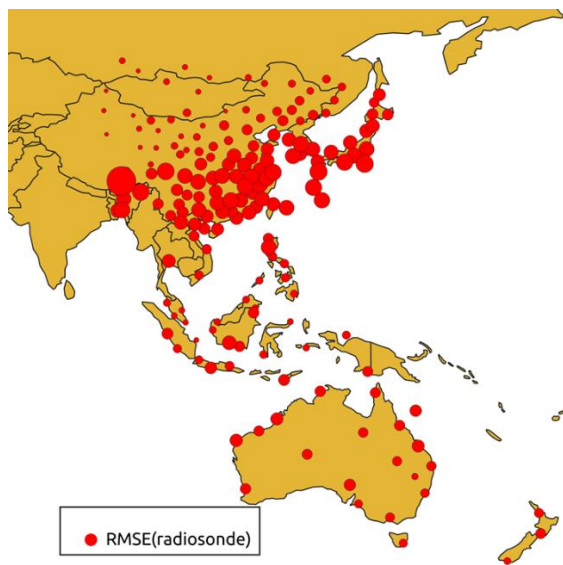


図-3 ラジオゾンデ観測点の位置と RMSE の関係
(RMSE(radiosonde)の大きさは RMSE に対応)

b) RMSE の空間的特徴の検討

各観測点における RMSE と位置の関係を調べた(図-3)。結果としては、内陸部において RMSE が小さくなる傾向が見られた。また、VZA および Ele の RMSE との関係を調べたが、相関は見られなかった。各月・各観測点における RMSE と各観測点の Lat, Ele, VZA との関係を調べた。その結果、全てにおいて相関は見られなかった。よって、観測点の緯度、標高および VZA 等の位置条件は推定精度に影響を及ぼさないと考えられる。

c) 月毎の重要変数についての検討

月毎に重要度の高い説明変数について調べた。結果として、年間を通して Ele が重要変数となったが、その他の重要変数は月毎に異なることが分かった。したがって、年間を通して 1 つの推定モデルで可降水量を推定するのは困難であり、月毎に可降水量推定モデルを作る必要があることが示唆された。

4. おわりに

本研究では、気象衛星ひまわりの観測データを用いた可降水量推定手法を、機械学習手法の一つであるランダムフォレストを用いて検討した。目的変数を可降水量、説明変数をラジオゾンデ観測点の緯度、標高、ひまわりの赤外バンドの輝度温度等とし、

Permutation Importance という測定方法で変数の重要度を示した。また、解析結果を月毎、観測点毎に比較することで推定モデルの汎用性を調べた結果、以下のことが分かった。2016 年データを基に作成された推定モデルを他の年へ適用することは困難であり、月毎に重要変数も異なるため年間を通して 1 つの推定モデルを適用させることもできない。また、推定モデルでは内陸部において RMSE が小さくなる傾向が見られたものの、それ以外の位置条件による推定精度の変化は見られなかった。

参考文献

- 1) 赤塚慎：気象衛星ひまわりによる可降水量推定手法の検討，日本写真測量学会平成 31 年度秋季学術講演会発表 論文集，2019
- 2) 赤塚慎，大吉慶，竹内渉：運輸多目的衛星 MTSAT 観測による可降水量推定手法の開発，日本リモートセンシング学会誌，Vol. 31, No. 5, pp.481-489, 2011
- 3) Julian Avila, Trent Hauck (株式会社クイープ訳)：Python 機械学習ライブラリ scikit-learn 活用レシピ 80+, pp. 279, インプレス
- 4) 志水菊広，齋藤幸太郎，山本幹人：ひまわり 8 号・AHI の 16 バンド画像の特性，気象衛星センター技術報告，第 62 号，2017 年 3 月
- 5) Molnar, Christoph: "Interpretable machine learning. A Guide for Making Black Box Models Explainable", 2019.
<https://christophm.github.io/interpretable-ml-book/>.
- 6) 樋口篤志・竹中栄晶・豊嶋絃一：ひまわり 8/9 号 フルディスク (FD) gridded data (緯度経度直交座標系精密幾何補正済データ) Version 02 (V20190123) 公開について，www.cr.chiba-u.jp/databases/GEO/H8_9/FD/index_jp.html
- 7) Weather Web. <http://weather.uwyo.edu/> (2019年9月17日アクセス)