

# 機械学習を用いた飛来塩分量推定手法の高精度化に関する研究

新潟大学大学院 学生会員 ○西村駿希  
新潟大学 正会員 佐伯竜彦  
新潟大学 正会員 齋藤豪

## 1. 研究目的

飛来塩分量は気象や周辺地形など様々な要因の影響を受け、定量的な評価が困難であるため、近年では機械学習を用いた推定手法が提案されている<sup>1)</sup>など。しかし、予測傾向に極端な偏りが確認されることや、特定の地域のみが対象となっているなど課題が残されている。

そこで本研究では、機械学習を用いた日本全国に適用可能な飛来塩分量推定手法の構築とその高精度化を目的として、MSM 風速データの精度に関する検討、降水データを追加の説明変数とした飛来塩分の沈降・洗い流しの影響評価、機械学習に使用する訓練データ数の影響の3点について検討を行った。

## 2. 使用データおよび処理

### 2.1 使用した飛来塩分量調査データ

本研究では、シリーズ I、II および飛来塩分量全国調査<sup>2)</sup>の3つの飛来塩分量調査結果を用いて検討を行った。測定地点はそれぞれ図1に示す通りである。

(1)シリーズ I : 新潟県沿岸部の 15 地点において、1992 年 9 月～1995 年 3 月の間で行われた調査データと、新潟県が 2002 年 3 月～2004 年 3 月に佐渡ヶ島にて実施した1カ月ごとの調査データを使用した。

(2)シリーズ II : 日本全国の計 101 地点で 2016 年 11 月～2017 年 4 月に実施された薄板モルタル供試体の暴露試験結果を使用した。

(3)飛来塩分量全国調査の結果<sup>2)</sup> : 飛来塩分量全国調査(III)に記載されている 1984 年 12 月～1987 年 11 月にかけて旧建設省によって全国で行われた土研式塩分捕集器による 1 カ月ごとの調査結果を使用した。ただし、当時と海岸形状が大きく変化していると考えられるもの、資料中に記載された地点が現在確認できないものは除外し、79 地点分のデータを使用した。

### 2.2 風況データの取得

対象とした飛来塩分量調査地点のそれぞれに対して、長期的な風況(風向,風速)の観測を行うことが困難であること、AMeDAS の風況観測地点が約 21 km 間隔であるため

に地点によっては観測地と距離が大きく離れ、風況が異なると考えられることから、本研究では MSM の風況データを用いた。

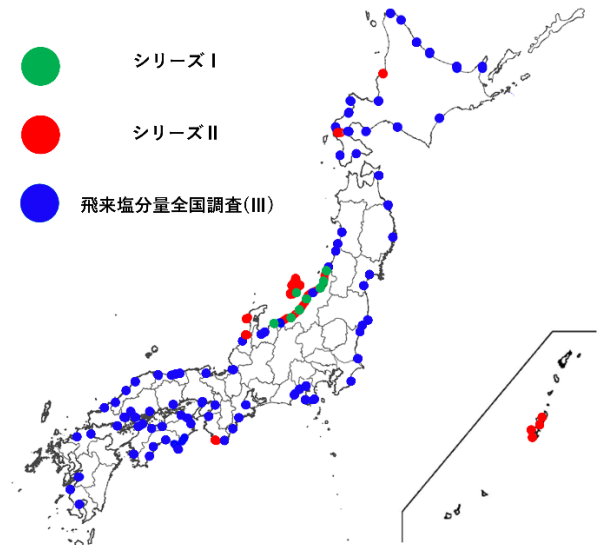


図1 飛来塩分量調査地点

### 2.3 検討対象とする風の範囲と補正

海風のみを対象としていること、飛来塩分の捕集に用いる装置(土研式塩分捕集器,薄板モルタル供試体)では、裏面方向の風によって輸送される飛来塩分を捕集することができないことから、飛来塩分の捕集に寄与しないと思われる範囲からの風の影響を取り除くために、検討に用いる風向の範囲を捕集口の向き $\pm 90^\circ$  以内かつ汀線水平方向までの範囲に制限した。田村らの報告<sup>1)</sup>に倣って、風速の閾値を $4\text{m/s}$ とし、式(1)によって風向による補正を行った。

$$v_i'^2 = v_i^2 \cos \theta_i \quad \dots(1)$$

ここで、

$v_i'^2$ : 角度補正した風速の二乗( $\text{m}^2/\text{s}^2$ )

$v_i^2$ : 1 時間ごとの平均風速の二乗( $\text{m}^2/\text{s}^2$ )

$\theta_i$ : 1 時間ごとの平均風向と塩分捕集器または供試体暴露面のなす角度

### 2.4 加重平均による MSM 風速の補正

MSM は 5 kmメッシュの格子点で風速が与えられているが、飛来塩分調査地点と格子点の位置が一致しているとは限らない。飯泉らは、対象とする地点の周囲の MSM 格

子点の値をその地点からの距離を用いた加重平均で補間することによって風速の実測値と推定値の誤差量が小さくなることを報告している<sup>3)</sup>。1 格子点のデータのみでは周辺環境の影響や風況の局所的变化を十分に考慮できないと考え、本研究においても目的地点の周囲 4 格子点のデータを使用して補正を行った。

$$U_H = \frac{\sum_{i=1}^4 \frac{1}{L_i} U_{MSMi}}{\sum_{i=1}^4 \frac{1}{L_i}} \quad \dots(2)$$

ここで、

$U_H$ : 飛来塩分量測定地点での風速

$U_{MSMi}$ : MSM 格子点での風速

$L_i$ : 飛来塩分量測定地点と格子点との距離

海岸沿いに位置する 40 地点の AMeDAS 観測点(図 2)の 2010 年から 10 年分のデータを使用した MSM データとの比較において、両者の回帰係数の Q-Q プロットは図 3 に示す通り直線で、アンダーソン・ダーリング検定量が  $P=0.45$  と  $0.05$  より大きくなったことから回帰係数の全国的なばらつきは正規分布に従うものと判断できる。有意水準 95%で回帰係数の区間推定を行った結果は 0.99~1.12 の間であり、少なくとも海岸沿いについては加重平均後の MSM 風速をそのまま使用して問題ないと考えられる。

調査地点から最短の距離にある 1 格子点の MSM データを使用した場合は、回帰係数の区間が 0.95~1.12 と加重平均を行うよりも幅が広く、加重平均による補正を行った方が高い精度の風速データを得られると考えられる。



図 2 比較に使用した地点の分布

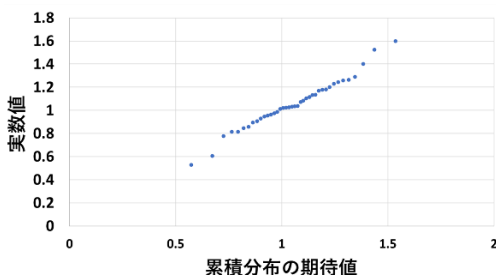


図 3 回帰係数の Q-Q プロット

## 2.5 地形データの取得と処理

飛来塩分量への影響が報告されている汀線から飛来塩分量測定地点までの距離<sup>4)</sup>、消波施設から飛来塩分量測定地点までの距離<sup>5)</sup>、障害物による遮蔽<sup>6)</sup>の値を GoogleEarth の距離測定機能を利用して求めた。汀線までの距離は、波跡の位置までで計測するものとし、干満による汀線位置の変化については、経時変化の把握が困難であることから考慮していない。機械学習を行う際に欠損値として扱われることを避けるために、周辺に消波施設が見られない場合の消波施設までの距離については無限遠方を仮定して 10,000,000(m)とした。遮蔽率は 2.3 で設定した計算対象範囲を占める住宅や樹木の割合から式(3)を使用して計算し、遮蔽している角度の大きさは GoogleEarth の距離測定機能から方位角の差によって求めた。

$$P = \frac{\sum \theta_i}{\phi} \quad \dots(3)$$

ここで、

$P$ : 遮蔽率

$\theta_i$ : 遮蔽されている角度

$\phi$ : 風速の検討範囲の角度

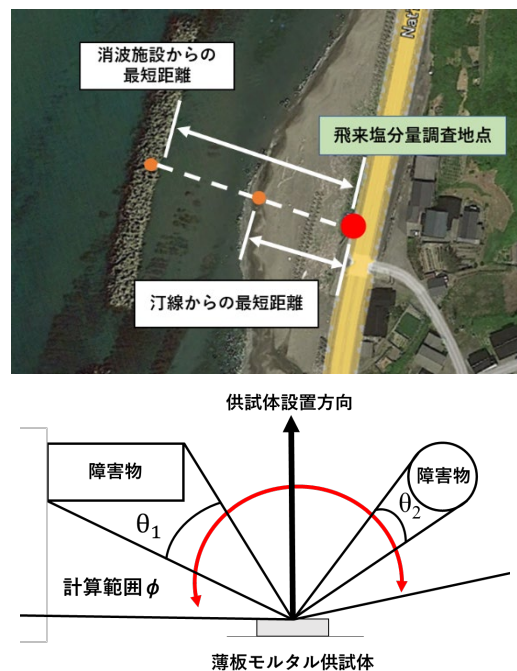


図 4 汀線からの距離、消波施設からの距離および遮蔽率の測定法

## 2.6 MSM データの取得ができない期間の風況データ

2006 年 3 月以前については現在と同精度の MSM データが提供されていないため、飛来塩分量データのうちシリーズ I の一部や飛来塩分量全国調査(III)の飛来塩分量測定期間に対応する気象データを取得することができない。そこで、式(4)によって推定値を算出した。

$$U_{ei} = \frac{U_{AME_i}}{\Sigma U_{AME}} \cdot \Sigma U_{MSM} \quad \dots(4)$$

ここで、

$U_{ei}$ : 調査当時の風速の二乗の積算値の推定値(m<sup>2</sup>/s<sup>2</sup>)

$U_{AME_i}$ : 調査地点に最も近い AMeDAS データを使用して求めた調査当時の風速の二乗の積算値(m<sup>2</sup>/s<sup>2</sup>)

$\Sigma U_{AME}$ : 調査地点に最も近い AMeDAS データを使用して求めた調査当時付近 10 年分の風速の二乗の積算値の平均値(m<sup>2</sup>/s<sup>2</sup>)

$\Sigma U_{MSM}$ : MSM データを使用して求めた直近 10 年間の風速の二乗の積算値の平均値(m<sup>2</sup>/s<sup>2</sup>)

検証として行った直近 10 年以前の風況の推定の例として、2010 年の風速の二乗の積算値の推定結果の例を図 5 に示す。風況データは図 2 に示す地点のものを使用した。MSM データと推定値に高い相関が得られたことから、推定値を使用した場合であっても、同年の MSM データを使用した場合と同等の結果が得られると考えられる。

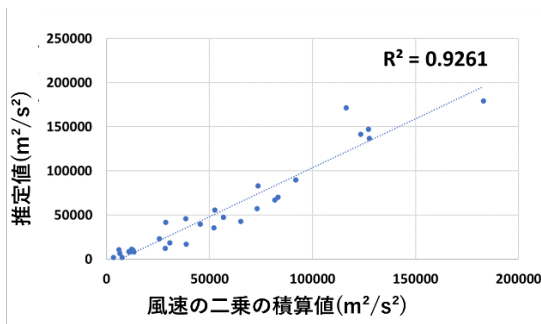


図 5 推定値と MSM データの相関

## 2.7 降水データの取得と処理

MSM データから取得した時間降水量を風速と同様に周囲 4 点の加重平均を求め、目的地点の時間降水量とした。2006 年 3 月以前のデータについても風速と同様に推定を行った。気象庁の定める用語を参考に、少雨 0.1, 0.5, 1(mm/h)と弱めの雨 3(mm/h), 5(mm/h)を閾値として、調査期間中の降水量の総和と、降水時間の総和を求めて検討を行った。5mm/h 以上の降水については非常に頻度が低く、説明変数として不十分であったために閾値としては使用していない。

## 2.8 使用した機械学習アルゴリズム

本研究では、不要な特徴量を入れても精度が落ちにくいこと、説明変数の標準化が必要ないこと等の理由から機械学習アルゴリズムに LightGBM を使用し、過学習をなるべく避けるため K-分割交差検証(分割数 K=10)を採用している。

## 3. 機械学習による推定の結果

### 3.1 使用データが学習精度に与える影響

シード値によりシリーズ I, II を学習させた結果(表 1)からは推定精度が学習データの並び順に大きく影響されていることが分かる。加えて、飛来塩分量全国調査を追加した際の学習結果では、全体的な推定精度が向上し、決定係数のばらつきも小さくなっている。このことから、推定精度の大きな変動は、データ数の不足が原因であり、今回の検討によって改善されたと考えられる。学習結果を確認すると、飛来塩分量 1000mg/dm<sup>2</sup>付近で推定誤差が大きくなっており、該当するデータが含まれる地点の塩分量データを確認すると、傾向を大きく外れて塩分量が高くなっている期間が見られた。このように同地点内で大きく外れる値を改めて異常値として除外したところ高い精度での推定が可能となった。

表 1 学習結果のまとめと学習順の影響

| シード値   | 決定係数(R <sup>2</sup> )<br>(シリーズ I, II) | 決定係数<br>(シリーズ I ~ III) | 決定係数<br>(異常値再除去) |
|--------|---------------------------------------|------------------------|------------------|
| 1      | 0.776                                 | 0.826                  | 0.921            |
| 2      | 0.115                                 | 0.840                  | 0.918            |
| 3      | 0.638                                 | 0.850                  | 0.919            |
| 4      | 0.686                                 | 0.842                  | 0.914            |
| 5      | 0.849                                 | 0.862                  | 0.932            |
| 6      | 0.954                                 | 0.827                  | 0.915            |
| 7      | 0.686                                 | 0.848                  | 0.896            |
| 8      | 0.754                                 | 0.843                  | 0.898            |
| 9      | 0.603                                 | 0.815                  | 0.951            |
| 10     | 0.752                                 | 0.828                  | 0.942            |
| 平均     | 0.681                                 | 0.838                  | 0.921            |
| 標本標準偏差 | 0.224                                 | 0.014                  | 0.017            |

### 3.2 風の加重平均

決定係数の平均値は加重平均による補正の前で 0.931、補正後で 0.921 となっており、殆ど差がみられなかった。LightGBM による学習では、データがヒストグラム化されて処理されるため、図 6 のように補正の前後で全体の分布傾向が似通っている場合、作成される決定木の分岐にも変化がなくなる。加重平均を取ることによって実測値との相関が上がっていたにも関わらず推定精度に変化が見られなかったのはこのことが原因と考えられ、決定木を使用した機械学習においては、風速の精度はある一定以上からは推定精度に大きく影響しない可能性がある。

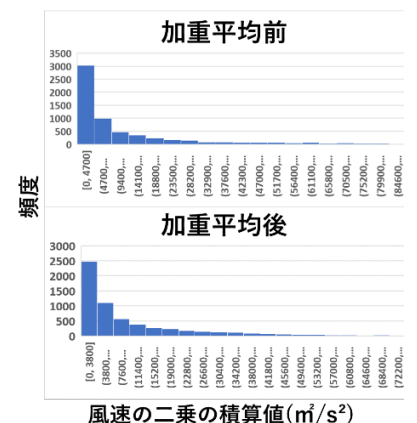


図 6 風速の積算値の頻度分布

### 3.3 降水の影響

降水量を降水時間で除し、調査期間内の降水時における平均降水強度に算出して特徴量として使用した。図 7 に示すように閾値に関わらず精度が向上している。図 8 に示した SHAP 値の傾向は降水強度が上がるほど塩分量が少なく推定されるものとなっており、妥当だと言える。いずれの閾値についても精度が変わらないため、少なくとも 5mm/h 以下の降水については影響量が小さいと考えられる。推定結果は図 9 に示す通りで、推定傾向に極端な偏りも確認されないため、日本全国の地点について実測値を概ね推定できている。

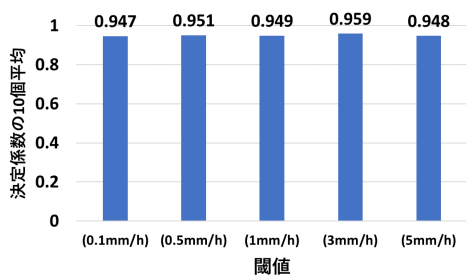


図 7 閾値ごとの学習精度

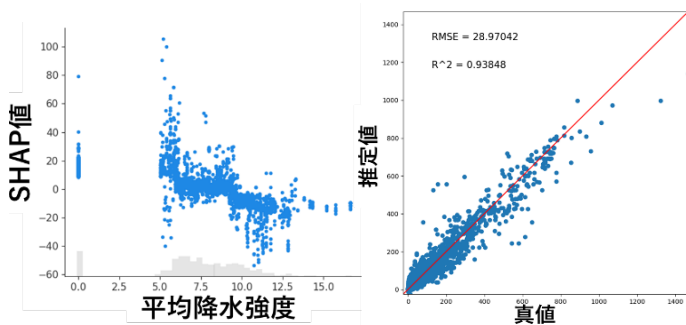


図 8 降水強度の SHAP 値の傾向(左)

図 9 飛来塩分量の推定結果(右)

### 3.4 飛来塩分量が未知である地点の推定

北西側と南西側で飛来塩分量に明確な差が出るため傾向の確認がしやすい佐渡ヶ島を対象として、実測値のない地点に対しても推定モデルが適用可能か検討を行った。気象データはシリーズⅡと同期間のものを使用した。推定にあたって学習に使用した地点を避けているため、厳密に同地点での比較とはなっていないが、中村らの現地観測結果<sup>4)</sup>と比較しても塩分量分布の傾向をよく再現できていると言える。

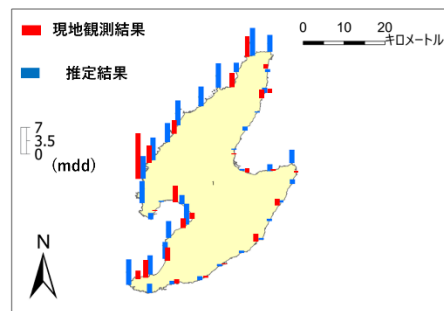


図 10 飛来塩分量の推定結果と現地観測結果<sup>4)</sup>

## 4. 結論

本研究では、気象データと周辺地形データを利用した飛来塩分量推定手法の高精度化を目的として、風速の補正、降水データの使用、学習データの追加を行い、各説明変数がどのように飛来塩分量に影響を及ぼすかの検討を行った。本研究で得られた結論は以下の通りである。

- 1). 降水強度を説明変数として加えることによって推定精度が向上する。
- 2). 機械学習を用いて構築した飛来塩分量推定モデルは、日本全国の地点について実測値を概ね推定できることを確認できた。
- 3). 機械学習を用いて構築した飛来塩分量推定モデルは実測値のない地点に対しても概ね実現象に沿った推定ができることを確認できた。

### 参考文献

- 1) 田村勇斗, 佐伯竜彦, 斎藤 豪: 風況と周辺地形を考慮した飛来塩分量予測手法に関する研究, コンクリート工学論文集, 第 33 巻, pp1-13, 2022
- 2) 土木研究所: 飛来塩分量全国調査(Ⅲ), 土木研究所資料第 2687 号, 1988.12.1
- 3) 飯泉元気, 谷口 円, 阿部 佑平, 堤 拓哉, 月館 司: 建築分野における技術開発等に利用する基礎的な気象データの構築, 地方独立行政法人北海道立総合研究機構建築研究本部北方建築総合研究所, 調査研究報告, No.416, 2021.3
- 4) 中村文則, Chimedtseren OTGONBILEG, 増田 健, 神田佳一, 下村 匠: 北陸・東北地方の日本沿岸部における飛来塩分の現地観測と環境作用の影響, 土木学会論文集 B2, Vol.77, No.2, I\_445-I\_450, 2021
- 5) 中村文則, 大原良平, 山口貴幸, 下村 匠: 沿岸域の地形条件を含めたコンクリート構造物の表面塩分の定量的評価手法の構築, 土木学会論文集 E2, Vol.76, No.3, pp189-208, 2020
- 6) 中村文則, 小林豪毅, 島田玄太, 永田千広, 高木利光, 児玉敏雄: 海岸近傍に設置した透過性構造物の飛来塩分の低減効果に関する現地実証実験, 土木学会論文集 B2, Vol.67, No.2, I\_1061-I\_1065, 2011