

# ランダムフォレスト回帰を利用した干拓地土壌の水分予測

## Prediction of soil moisture in reclaimed land using Random forest regression

升田直希<sup>1</sup>・平嶋雄太<sup>2</sup>・宮本英揮<sup>2</sup>

<sup>1</sup>佐賀大学大学院農学研究科・<sup>2</sup>佐賀大学農学部

### 要旨(Abstract) :

干拓地の農地環境データに基づき、ランダムフォレスト(RF)回帰アルゴリズムを活用した誘電率( $\epsilon$ )予測モデルを構築し、6時間後の表層土中の $\epsilon$ の予測を行った。降雨開始直後の数時間を除けば、比較的良好な精度で $\epsilon$ の予測値を得られることが明らかとなった。

キーワード：機械学習、ランダムフォレスト回帰、土壌水分量、干拓地

Key words: Machine learning, Random forest regression, Soil moisture, Reclaimed land

### 1. はじめに

一般に、下層の高塩分濃度層からの塩分の遡上のリスクを抱える海成干拓地では、塩分とその輸送に関わる土中水の動態を把握し、それらを制御することが重要である。IoT技術の発展に伴いビッグデータの収集が可能となった昨今、データ駆動型の土壌水分量・塩分濃度予測が可能になりつつある。本研究では、ランダムフォレスト回帰による表層土の水分予測を試み、その有効性を検討した。

### 2. 方法

長崎県農林技術開発センター・干拓営農研究室の試験圃場の-5 cmに水平に埋設したTDR-315L センサ(Acclima)を用いて見かけの誘電率( $\epsilon$ )を、また、同一圃場内の気象観測システムを用いて気温( $T$ )、相対湿度( $RH$ )、風速( $WS$ )、日射量( $I$ )、降水量( $Pr$ )を、それぞれ1時間間隔で測定した。これらのデータに基づく繰り返し計算法<sup>1)</sup>により、降水量に基づき実効雨量( $D$ )<sup>2)</sup>を算出した。先行研究<sup>3)</sup>を踏まえ、2018年8月15日～2020年12月31日のデータを $\epsilon$ 予測モデル構築のための学習用データ、2021年1月1日～同年12月31日のデータを検証用データとそれぞれ定めた。

予測実施時刻を起点( $K = P = 0$ )とし、 $P$ (本研究では $P = 6$ )時間後の $\epsilon$ を目的変数、 $K$ ( $K = 0, 2, 5, 11, 17, 23, 35, 47, 71, 95, 119$ )時間前までの $D, T, RH, WS, I, DOY$ をそれぞれ説明変数と定めたうえで(Fig.1)、計11種の $\epsilon$ 予測モデルを構築した。Python言語用のオープンソース機械学習ライブラ

リ・scikit-learnを利用したランダムフォレスト(以降、RFと称す)回帰による $\epsilon$ 予測モデルを、次の手順で構築した。まず、学習用データからランダムに $\epsilon$ および各説明変数のデータセットを抽出した。次に、 $\epsilon$ の平均値に対する $\epsilon$ の残差平方和が最小となるように説明変数の閾値を設け、それを境に $\epsilon$ のデータセットを2分割した。この操作を2分割後の各データセットにおいても実施し、さらに分割後の $\epsilon$ のデータ数が1つになるまで反復することで、階層構造を有する回帰木<sup>4)</sup>を定めた(Fig.1)。一連の手順を100反復し、得られた計100種の回帰木(Tree1～Tree100)に検証用データの説明変数を代入することにより得られた $\epsilon_1 \sim \epsilon_{100}$ の平均値を求めた。ランダムにデータを抽出するRF回帰では、先述の予測を行うたびに異なる予測値が得られる。本研究では、各条件において予測値を3回算出し、それらの平均値を $\epsilon$ の予測値と定めた。検証用データに対する予測値の $R^2$ およびRMSEに基づ

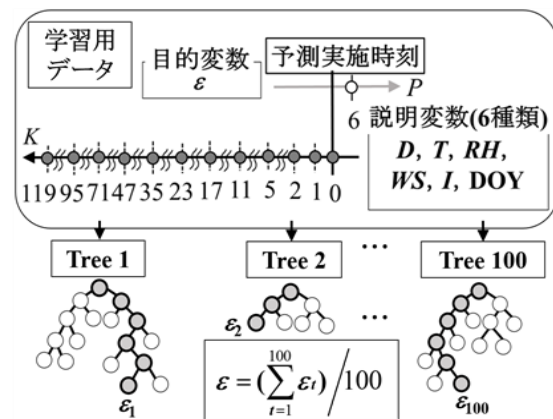


Fig.1  $\epsilon$ 予測モデルの構築イメージ

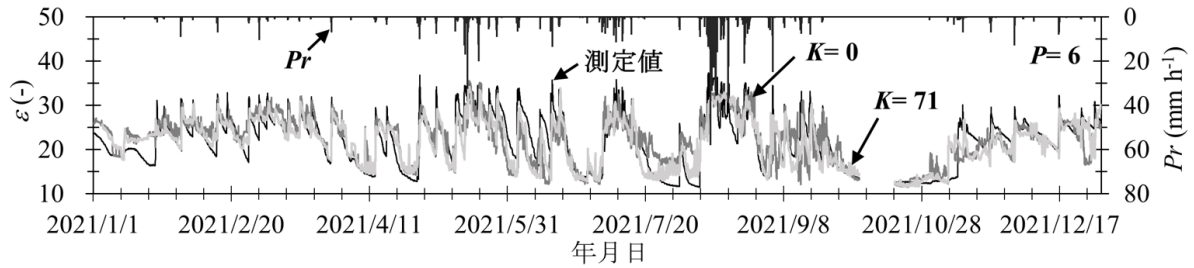


Fig.2 -5 cm の誘電率( $\epsilon$ )の測定値およびその予測値と降水量( $Pr$ )の変化

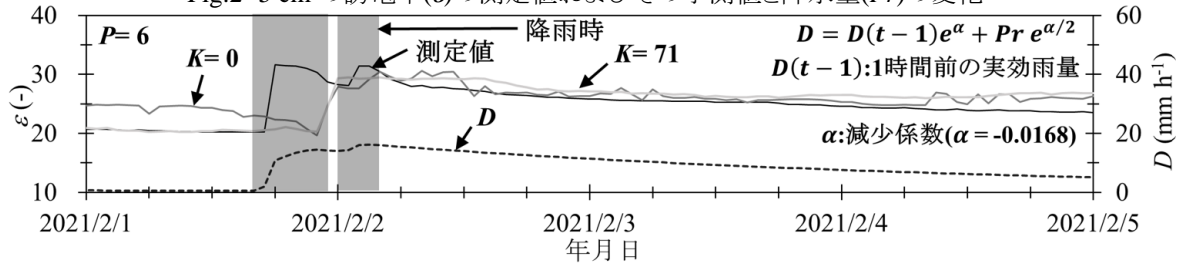


Fig.3 -5 cm の誘電率( $\epsilon$ )の測定値およびその予測値と実効雨量( $D$ )の変化

き,  $\epsilon$ 予測モデルの精度を評価した。

### 3. 結果と考察

降雨開始直後の数時間を除き,  $\epsilon$ の予測値は測定値と連動した。降雨時に測定値は上昇し, 無降雨時に低下する一般的な変化が認められた (Fig.2)。一方,  $\epsilon$ の予測値の一例として示した, 予測実施時刻( $K=0$ )の気象データを説明変数とした場合の6時間後( $P=6$ )の予測値は, 降雨時に上昇し, 無降雨時に低下した (Fig.2)。類似した予測値の変化は, 全  $K$  条件に共通して認められ (図は割愛),  $P=6$  では  $K=71$ , すなわち, 72時間分の説明変数に基づく予測モデルの精度が最良であった (Fig.4)。

降雨開始直後の変化に着目すると, 測定値が上昇した数時間後に予測値が上昇した。測定値は,  $D$  と概ね連動した上昇傾向を示したが (Fig.3), 予測値は  $D$  の上昇開始から約6時間後 (例えば, 2/1 23:00) に上昇した (Fig.3)。予測値の上昇開始時刻の遅延は, 予測対象時刻の直前の  $P$  時間分の  $D$  を説明変数として予測モデルに組み込むことができず, 直近の降雨を予測結果に反映できないためである。類似した予測値の上昇開始時刻の遅延は, 降雨時に共通する事象として全期間に認められた (図は割愛)。

無降雨時に限定すれば, 測定値に近い予測結果が得られた。予測値の上昇の遅延が認められた降雨時と違い, 無降雨時の予測値は測定値と概ね同様の減少傾向を示し (Fig.3), それは全期

● :RMSE(全期間)	○ : $R^2$ (全期間)
▲ :RMSE(降雨時)	△ : $R^2$ (降雨時)
■ :RMSE(無降雨時)	□ : $R^2$ (無降雨時)

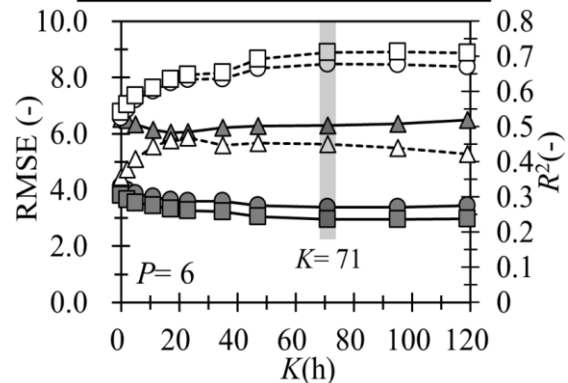


Fig.4  $K$ と予測精度との関係

間を通して確認された (図は割愛)。降雨時および無降雨時の  $R^2$  および RMSE に基づき予測精度を比較した結果, 全  $K$  条件に共通して, 無降雨時の予測精度が高いことが判明した (Fig.4)。

### 5. おわりに

降雨直後の数時間を除けば, RF 回帰を利用して, 6時間後の  $\epsilon$  を比較的良好な精度で予測可能であることが明らかになった。今後は, 異なる  $P$  条件における予測結果を検証し, RF 回帰による予測モデルの有効性と限界を評価する予定である。

謝辞: 長崎県農林技術開発センター・干拓営農研究室の助力を得て本研究を実施した。ここに記して謝意を表す。

参考文献: 1) 升田ら(2022): 農業農村工学会九州沖縄支部大会講演要旨集, 164-165., 2) 鈴木・小橋(1981): 砂防学会誌, 34(2): 16 - 24., 3) 平嶋ら(2020): 農業農村工学会論文集, 88(1): I\_21-I\_28., 4) Trevor et al(2014): 統計的学習の基礎 データマイニング・推論・予測, 共立出版。