

コロンビア水田における土壌の乾燥過程に関するデータ駆動型モデリング Data-driven modelling of spatiotemporal soil-water dynamics in paddy fields in Colombia

○福田信二¹・高橋太郎^{2,3}・Lorena López-Galvis³・Dario Pineda⁴

○FUKUDA Shinji, TAKAHASHI Taro, LÓPEZ-GALVIS Lorena, PINEDA Dario

1. はじめに

イネは世界三大作物の一つであり、コメは世界人口の半数以上にとって主食となっている。コメ生産量の90%以上はアジアで生産されているが、ここ数十年間の生産量の伸びは中南米やアフリカで大きく、一人当たりのコメ消費量もラテンアメリカとアフリカで大きく増加している。ラテンアメリカの稲作はアジアとは異なり、直播と大型機械の利用による大規模生産体系である。水管理は、等高線に沿ったタイパと呼ばれる低い畔を用いた掛け流しの田越灌漑がほとんどであるため、水および肥料の利用効率が非常に低い。そのため、適切な水管理と省資源型稲作を可能にする新品種およびその能力を最大限に活かす管理手法等の開発が強く望まれている。そこで本報では、コロンビア国イバゲ県の研究対象圃場における土壌の乾燥過程の連続観測および機械学習を用いた土壌水分の時空間変動の解析結果について報告し、水文研究におけるデータ駆動型モデル（data-driven model）の適用可能性について議論する。

2. 方法

水田土壌の乾燥過程のモデル化のために、コロンビア国イバゲ県の研究対象圃場（Potrerito, Lapilar, Perales）において、土壌水分の集中観測を実施した（Fig. 1）。まず、灌漑により土壌を飽和状態にし、12日間の無降雨期間中に毎日TDR土壌水分センサー（HydroSense II, Campbell Scientific, Inc., USA；ロッド長12 cmおよび20 cm）を用いて体積含水率（%）を計測した。その際、等高線を基準に設定した測線に従って計測し、圃場内の水分状態の不均一性の評価を試みた。本報では、圃場内の土壌水分の時空間変動の解析手法として、ランダムフォレスト

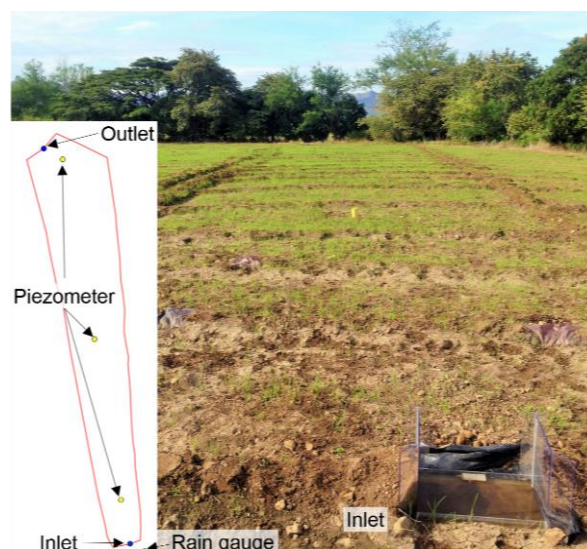


Fig. 1 Experimental setting at the Potrerito station

（Breiman, 2001）を適用した。解析には、統計ソフトR（R Core Team, 2015）のパッケージ“randomForest”（Liaw and Wiener, 2002）を使用し、変数の重要度および応答曲線に基づいて、乾燥過程を支配する要因や環境条件を評価した。再現性の評価には、平均二乗誤差（MSE）およびNash-Sutcliffe Efficiency（NSE）を使用し、50回の独立した試行によって、モデルの構造のばらつきを評価した。

3. 結果と考察

ランダムフォレストにより、異なる深度（12cmと20cm）における平均的な土壌水分量の時空

¹ 東京農工大学大学院農学研究科 Institute of Agriculture, Tokyo University of Agriculture and Technology

² School of Veterinary Sciences, University of Bristol

³ 東京大学大学院農学生命科学研究科 Graduate School of Agricultural and Life Sciences, The University of Tokyo

⁴ FEDEARROZ, Federación de Productores de Arroz, Colombia

間変動を比較的良好な精度で再現することができた (Fig. 2, Table 1) . ロッド長12cmのセンサーでの調査については、畔部 (caballone) と平地部 (melga) を区別して測定したため、調査地点数および解析に使用した変数の個数が異なるが、同データで構築したモデル (RF_{12cm}) の方が高い再現性を示した (Table 1) . 変数の重要度から、気象条件よりも位置に関する情報 (特に、畔部の考慮) が重要であることが示唆された (Fig. 3) . 応答曲線から実験初期に土壌水分が急激に減少したことや畔部において乾燥が進んでいること、最大風速4.5 m/s 以上では土壌水分がほぼ一定の応答を示す等の傾向がみられた (Fig. 4) . ただし、データ駆動型モデルは物理プロセスを反映しているわけではないため、結果の解釈には注意が必要である.

以上のように、対象とする現象の再現に重要な環境変数の特定や環境条件の数理解析によって、対象システムに内在する非線形的で複雑な関係性が観測データを通して可視化できるようになる. これにより、現象の理解が深化し、解釈に注意を要するものの、プロセスベースモデルへのフィードバックも可能であると考えられる.

4. おわりに

本報では、水田における土壌水分の時空間変動解析に対するデータ駆動モデルの適用例を紹介した. 一般に、流域水環境の解析には、研究対象と解析スケールに応じて必要な情報量が決定づけられ、精緻なモデルほど大量かつ詳細なデータが要求される. 一方、環境モニタリングの重要性が広く認識されたことにより、多様な機関によって水環境データを収集されており、日本国内においてもオープンデータとして公開される情報が増加している. そのため、単に各種水環境情報の予測や推定が目的であれば、データ駆動型モデルは再現精度の面でも有効である. 比較的長い歴史を有するハイドロインフォマティクスの知識と経験をフル活用することにより、水環境ビッグデータに基づいた対象システムの理解深化や将来シナリオの策定等は十分に可能であると考えられる.

参考文献

- Breiman, L. (2001): Random forests. *Machine Learning* 45, 5–32.
- Liaw, A. and Wiener, M. (2002): Classification and regression by random forest. *R News* 2(3), 18–22.
- R Core Team (2015): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Table 1 Model performance.

Model	MSE	NSE
RF _{12cm}	6.017±0.039	0.890±0.001
RF _{20cm}	12.723±0.064	0.694±0.002

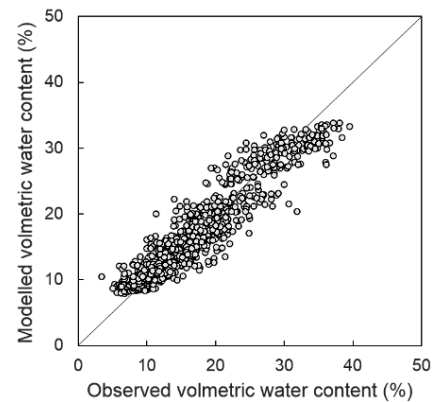


Fig. 2 Scatter plot between observed and modelled soil moisture content (RF_{12cm})

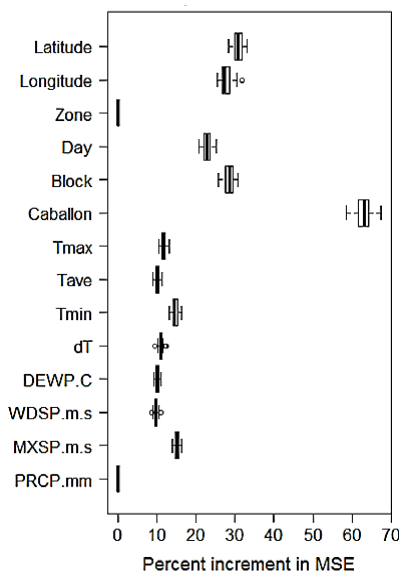


Fig. 3 Variable importance computed by RF_{12cm}

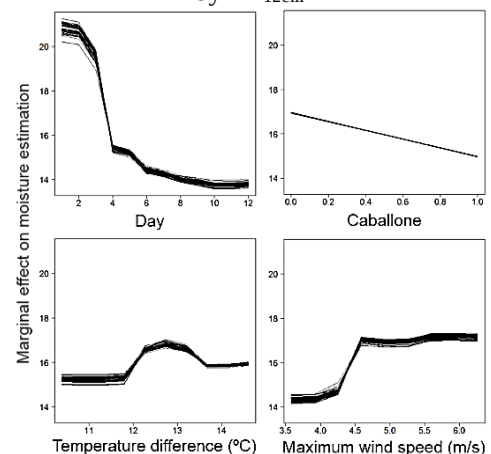


Fig. 4 Some response curves obtained from RF_{12cm}